When gender discrimination is not about gender

Katherine B. Coffman,
Christine L. Exley,
Muriel Niederle *†

November 8, 2018

Abstract

We use an experiment to show that employers prefer to hire male over female workers for a male-typed task even when they have identical resumes. Using a novel control condition, we document that this discrimination is not specific to gender. Employers are simply less willing to hire a worker from a group that performs worse on average, even when this group is instead defined by a non-stereotypical characteristic. Thus, our evidence points to an important role for beliefs in explaining gender discrimination. We also document evidence for in-group preferences that contribute to the gender discrimination observed. Finally, our design allows us to address the role of image concerns in driving our results.

1 Introduction

Understanding the drivers of gender differences in labor market outcomes has been an important topic of study among labor economists, with research identifying sizable roles for occupational segregation, differences in human capital accumulation, demand for flexibility, and differences in preferences (Goldin, 2014; Card, Cardoso and Kline, 2016; Olivetti and Petrongolo, 2016).[1]

Discrimination also contributes to these gender gaps in earnings and advancement. A large body of empirical work has documented the existence of gender discrimination in labor markets, often through clever field experiments or with quasi-experimental data (for reviews, see Riach

*Coffman: kcoffman@hbs.edu, Harvard Business School; Exley: clexley@hbs.edu, Harvard Business School; Niederle: niederle@stanford.edu, Stanford University, SIEPR and NBER.

†Thank you to John Beshears, Alison Wood Brooks, Lucas Coffman, Benjamin Edelman, Leslie John, Johanna Mollerstrom, Joshua Schwartzstein, Andrei Shleifer, and seminar audiences at Cornell University, the Rotman School of Management, and the 2016 ESA North American Meetings for their many useful comments.

[1]For reviews on gender differences, see Croson and Gneezy (2009), Bertrand (2011), Azmat and Petrongolo (2014) and Niederle (2016).

and Rich (2002) and Blau and Kahn (2017)). Discrimination against women has been found in bargaining contexts (Ayres and Siegelman, 1995; Castillo et al., 2013; Bowles, Babcock and Lai, 2007), in hiring, employment, and referral contexts (Neumark, Bank and Nort, 1996; Goldin and Rouse, 2000; Black and Strahan, 2001; Bertrand and Mullainathan, 2004; Moss-Racusin et al., 2012; Reuben, Sapienza and Zingales, 2014; Baert, De Pauw and Deschacht, 2016; Bohnet, van Geen and Bazerman, 2016; Sarsons, 2017*a*), and in academic contexts (Milkman, Akinola and Chugh, 2012, 2015; Sarsons, 2017*b*).[2]

Despite this large literature, important open questions about gender discrimination remain. One challenge is identifying the primary driver of observed discrimination. Our study attempts to answer this key question. Traditionally, economists have viewed discrimination through two distinct lenses: taste-based or statistical. The taste-based argument posits that discrimination is rooted in preferences, driven by animus or prejudice (Becker, 1957). Statistical discrimination, on the other hand, is rooted in rational beliefs about average gender differences in abilities or skills (Phelps, 1972; Arrow, 1973).[3] Recently, researchers have proposed more general forms of belief-based discrimination, with a particular focus on discrimination based upon inaccurate beliefs. Bordalo et al. (2016) document a role for representativeness-based stereotypes, where beliefs about groups are not accurate but instead are biased by differences in low probability but highly representative tails. These inaccurate beliefs can then give rise to self-stereotyping and discriminatory behavior (Coffman, 2014; Bordalo and Shleifer, 2018). In this spirit, Bohren, Imas and Rosenberg (2017) document discrimination against women on an online math platform and show that it is most consistent with discrimination based upon biased beliefs.

However, cleanly disentangling these different drivers of discrimination — particularly accurate beliefs, biased beliefs, and tastes — is challenging. Past literature has mainly followed three approaches. One approach is to test for a role for beliefs by increasing the amount of information available on individual workers. The argument is that if more informative performance information reduces discrimination, beliefs must play an important role. However, to the extent that discrimination remains, it is unclear whether it reflects taste-based considerations or residual belief-based motivations. A second approach is to simply remove information that identifies the (gender, ethnic, socio-economic, etc.) group to which an individual belongs, for instance by taking a name off of a resume, ensuring that any discrimination does not reflect taste-based motivations.[4] However, in contexts where there are group differences in performance

[2]Many scholars have identified discrimination against other groups as well, such as the beauty premium (Mobius and Rosenblat, 2006), discrimination based on socioeconomic status (Rao, 2018), and racial or ethnic discrimination (for examples, see Fershtman and Gneezy (2001), List (2004), Charles and Guryan (2008), Castillo and Petrie (2010), Gneezy (2012), Doleac and Stein (2013), Lei and Babcock (2013), and Bartoš et al. (2016), and, for a review, see Charles and Guryan (2011)). For a review, see also Bertrand and Duflo (2017).

[3]See Altonji and Blank (1999) and Guryan and Charles (2013) for review and discussion of literature on taste-based versus statistical discrimination.

[4]A distinct but related approach is to simulate a context where taste-based considerations are absent, for instance by studying anonymous workers with randomly-assigned productivities, allowing for clean documentation

on average, removing information that identifies the group to which an individual belongs not only rules out taste-based motivations, but also inhibits belief-based discrimination. A final approach, often used in field data, is to ask whether the observed discrimination is consistent with profit-maximization, suggesting statistical motivations. However, this has typically required assuming that the discrimination is driven by accurate, rather than inaccurate, beliefs.

Ideally, what one would do to separate any role of beliefs from tastes would be to recreate the exact hiring environment where we see discrimination against women, divorced from any gender-specific considerations but not ability-specific considerations. This is our goal. We employ a novel control treatment that allows us to benchmark the level of discrimination we observe against women in a male-typed hiring environment to discrimination against workers of unknown gender known to be drawn from the same ability distribution. We do this in a setting where we can provide accurate information both on individual workers and group differences and where we can measure employer beliefs.

In our first experiment, we collect performance information on easy and hard math and sports quizzes. We use these participants as available "workers" for hire in the second experiment. In our second experiment, we ask "employers" to make incentivized choices over available workers. Employers receive information about the easy quiz performances of available workers. When they choose to hire a worker, they are paid based upon the hard round performance of that worker, which is unknown to them at the time of the decision.

The second experiment involves two treatments that vary only in their labeling of the available workers: a Gender treatment and a Birth Month treatment. In the Gender treatment, participants make hiring decisions between female and male workers, labeled as such, while in the Birth Month treatment, participants make decisions over workers born in an even month or workers born in an odd month. Importantly, the actual set of available workers is held constant across the two treatments; we simply vary how the workers are labeled. To implement this, we use two subsamples to generate our pool of available workers: women born in even months and men born in odd months. Thus, while the labels used to describe the available workers vary by treatment, the performances of the available workers do not. In both treatments, participants receive accurate, detailed information about the distribution of performances across the groups prior to making their hiring decisions, generating similar ex-post beliefs about the average ability gap (male/female gap and odd/even gap) across the two treatments. In this way, we abstract away from the belief formation process and rely on decision-makers who, by design, hold detailed, approximately accurate beliefs about average performance differences.[5] This allows us to directly

of statistical discrimination (for instance, Dickinson and Oaxaca (2009) and Dickinson and Oaxaca (2014)). Of course, absent a comparable context where taste-based considerations *are* relevant, this approach does not allow for one to back out how much of the observed discrimination is attributable to beliefs versus tastes.

[5]This of course limits our ability to test the stereotype formation channel of Bordalo et al. (2016). We focus instead on decision-making given a certain set of beliefs.

explore how pivotal those beliefs are for discrimination.

Our new methodology has a number of advantages. First, we elicit employer beliefs of gender differences, allowing us to directly link beliefs to behavior. Second, our Birth Month treatment better isolates the gender-specific taste component of discrimination by holding beliefs, both about individuals and groups, constant. It does so by removing information about worker gender while simultaneously maintaining identical information about individual and group-level performances. In this way, it provides a useful benchmark to which we can compare the extent of discrimination against women in the Gender treatment. Holding fixed beliefs about individual performances and average performance differences across the groups, we can ask whether women are any less likely to be hired when they are labeled as women rather than as even-month workers. Third, by replacing the gender labels in the Gender treatment with birth month labels in the Birth Month treatment, we are able to speak to the relevance of in-group preference in hiring decisions, and whether such in-group preference, are gender-specific.

We find that, on average, women in the Gender treatment are less likely to be hired than equally able men. Perhaps surprisingly, however, this result is not specific to gender. When we turn attention to the Birth Month treatment, we observe very similar levels of discrimination against even-month workers. This suggests that behavior is not driven by animus toward women and instead is more consistent with belief-based theories of discrimination.

We also find heterogeneous treatment findings that are consistent with our results reflecting an in-group preference, although perhaps surprisingly again, one that is not specific to gender. Not only do female, as compared to male, employers hire female workers more often, even-month, as compared to odd-month, employers hire even-month workers more often.

Finally, we explore whether our results persist when there is an additional veil on employers' intentions, reducing image concerns. We operationalize this veil in a manner that is similar in spirit to the examination of risk as a veil in Exley (2015).[6] We continue to find that women are hired no less often when labeled as women than when labeled as even-month, suggesting that our failure to document clear evidence of taste-based discrimination against women is not driven simply by image concerns. The introduction of a veil of intentions does, however, eliminate in-group preferences, suggesting a significant role for image concerns in driving these findings.

2 Design

Our experiments are conducted on Amazon Mechanical Turk. In the first experiment, we collect performance information on easy and hard math and sports quizzes. We recruit 100 individuals for an advertised 40-minute academic study with a guaranteed completion fee of $6. The workers completed a series of six quizzes: two quizzes involving easy questions about sports, one quiz involving hard questions about sports, two quizzes involving easy questions

[6]See also Dana, Weber and Kuang (2007), Haisley and Weber (2010), Danilov and Saccardo (2017), and for reviews, Kunda (1990) and Gino, Norton and Weber (2016).

about math, and one quiz involving hard questions about math. Each quiz contained 10 multiple-choice questions; workers had three minutes per quiz to answer as many questions as possible. To incentivize performance, workers received, as bonus payment within one week, 10 cents for each question they answered correctly, for one randomly-selected quiz. The workers were also aware that their performances would potentially be shown to other participants in follow-up experiments. We use these participants as available "workers" for hire in the second experiment.

In the second experiment, we ask "employers" to make incentivized choices over available workers. We recruit 800 new Amazon Mechanical Turk participants for a 25-minute academic study with a guaranteed completion fee of $4. Employers receive information about the easy quiz performances of available workers. Then, they make a series of hiring decisions over a subset of the worker pool. To incentivize decisions, employers were aware that for one randomly-chosen hiring decision they would receive, as bonus payment within one week, any associated payoffs.

2.1 Treatment variations

Employers in our study make hiring decisions that involve two types of workers: (1) male workers born in odd months, and (2) female workers born in even months. Employers are not given the opportunity to hire any other workers (i.e., any male workers born in even months) from our worker pool. For clarity, we will refer to the groups of available workers as male-odd-month workers and female-even-month workers.

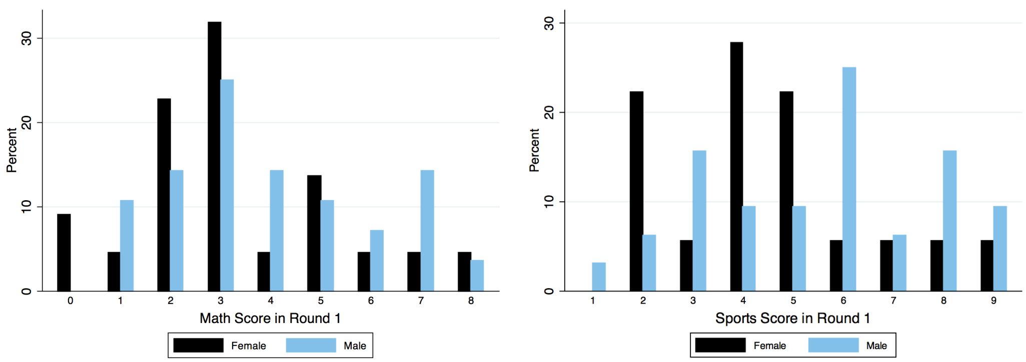
While employers always make decisions between male-odd-month workers and female-even-month workers, employers view different labels for these groups of workers. Employers in the Birth Month treatment are not provided with the gender labels, and thus see all information and decisions as between odd-month workers and even-month workers. Employers in the Gender treatment are not provided with the birth month labels, and thus see all information and decisions as between male workers and female workers. The treatments are identical outside of the labels provided. In addition, employers are randomly assigned to either the math or sports version of the study. All results below pool these two versions.

2.2 Information Stage

In order for the Birth Month treatment to serve as an interesting comparison treatment, employers must have similar beliefs about the performances of the two groups of workers in both treatments. We achieve this by providing accurate and comprehensive information: we show each employer the full distribution of performances for female-even-month workers and male-odd-month workers. In addition, we reinforce this information and increase the employers' engagement with it by requiring employers to make *incentivized* decisions that relate to the distributions of performances for various subsets of female-even-month workers and male-odd-month workers. All together, employers view 12 sets of distributions that detail the performance of workers on the easy quiz.

The first 11 sets of distributions are formed by restricting the subset of workers to those born during different date ranges (see Appendix Table A.1 for more details on how these distributions are formed). The 11 distributions are unique but overlapping (that is, an employer only sees the distributions for workers born between the 1^{st} - 15^{th} of the month once, but she also sees the distributions for workers born between the 1^{st} - 10^{th}). The order in which these 11 sets of distributions are shown is randomized at the participant level. For all participants, the 12^{th} and final set of distributions contains the full distributions: histograms of the performances of all male-odd-month workers and female-even-month workers, see Figure 1.

Figure 1: Worker Performance on the Easy quiz



(a) Math Version

(b) Sports Version

Note: Panel A shows the performance distributions in the math quiz for female-even-month workers, who have an average performance of 3.27, and male-odd-month workers, who have an average performance of 3.96. Panel B shows the performance distributions in the sports quiz for female-even-month workers, who have an average performance of 4.50, and male-odd-month workers, who have an average performance of 5.50. The labels "Female" and "Male" shown here are those that would appear in the Gender treatment. In the Birth Month treatment, the labels would instead read "Even" and "Odd", respectively.

We incentivize decisions that relate to each set of distributions as follows. Below each set of distributions, the employers are asked to select one group from which they would like to hire. In the Birth Month treatment, the options are "The Even Month Group", "The Odd Month Group" and "Let Chance Determine the Group." In the Gender treatment, the options are "The Female Group", "The Male Group" and "Let Chance Determine the Group." "Let Chance Determine the Group" results in a 50% chance of each group. If one of these decisions is randomly selected as the decision that counts for payment, one worker from the selected group is chosen at random to be hired. The hired worker receives an additional 25 cents as bonus payment. Employers receive 10 cents for each question answered correctly by the hired worked on the hard quiz.

2.3 Belief Elicitation Stages

Before completing the *Information Stage*, we elicit employers' prior beliefs about the performance gap between male-odd-month workers and female-even-month workers. Employers in the Birth Month (Gender) treatment are asked about their prior beliefs about the performance gap between odd-month (male) and even-month (female) workers. In particular, employers predict the difference in average scores between these groups of workers on the easy quiz and the hard quiz. Employers could indicate any feasible average difference (from -10 to 10 problems solved correctly), with positive numbers indicating a performance gap in favor of odd-month (male) workers in the Birth Month (Gender) treatment.

After completing the *Information Stage*, employers are asked the same beliefs questions again to determine their posterior beliefs. Neither the prior nor posterior beliefs are incentivized.

2.4 Hiring Stage

After the *Information Stage* and *Belief Elicitation Stages*, we ask employers to make a series of hiring decisions between specific pairs of workers. In each hiring decision, employers in the Birth Month treatment are asked whether they want to hire "The Even-Month Worker," the "Odd-Month Worker," or "Let Chance Determine Who is Hired." Employers in the Gender treatment are asked whether they want to hire the "The Female Worker," "The Male Worker," or "Let Chance Determine Who is Hired." For each hiring decision, employers learn the two workers' exact performances on the easy quiz. We vary the performances of the available workers and the payoffs to hiring each across a series of 54 decisions, split across six screens (within-subject).

On each screen in the *Hiring Stage*, employers make nine hiring decisions. For each, they are told the performances of the two workers in the pair. In three of the hiring decisions, the available workers have the same performance on the easy quiz. The number of questions answered correctly (out of 10) by the male-odd-month worker versus the female-even-month worker is (i) 4 versus 4, (ii) 6 versus 6, and (iii) 8 versus 8. In the other six hiring decisions on the screen, the male-odd-month worker has a weaker performance than the female-even-month worker: (iv) 4 versus 5, (v) 4 versus 6, (vi) 4 versus 7, (vii) 4 versus 8, (viii) 6 versus 7, and (ix) 6 versus 8.

Across the screens in the *Hiring Stage*, we vary the payoffs to hiring each worker. On the first *Hiring Stage* screen, the payoff to hiring either worker is 10 cents for each question answered correctly by their hired worker on the hard quiz.

In these stark, side-by-side hiring decisions, employers may feel that choosing not to hire the female-even-month worker (who always has a weakly better performance) is harmful to their self-image or social-image. This might be particularly true in the Gender treatment due to concerns about perceived sexism. So, similar in spirit to how Exley (2015) introduces risk to allow individuals to justify not donating money, we include additional "risk" treatments that provide a plausible alternative explanation for why an employer would not hire the female-even-month worker.[7] In these decisions, the payoff remains 10 cents per correct answer *if* the employer hires a male-odd-month worker but now involves risk if an employer hires a female-even-month worker. In particular, if an employer hires a female-even-month worker, she receives 10 cents for each question correctly answered by that worker on the hard quiz with P% chance but no payment with (1-P)% chance. Across the second through sixth *Hiring Stage* screen, P decreases from 99, 95, 90, 75 to 50. While the introduction of risk decreases the expected payment from hiring a female-even-month worker, we note that this decrease is particularly negligible when it implies a mere 1% chance of no payment and that our examination of hiring decisions under scenarios with risk does not result in different conclusions across each of these P values.

If one of these 54 hiring decisions is randomly selected to count for payment, employers receive the amount of money earned in the selected decision as an additional bonus payment. Their hired worker from the selected decision also receives 25 cents as additional bonus payment.

3 Results

We begin by examining employers' beliefs in Section 3.1. We then investigate whether discrimination influences overall hiring decisions in Section 3.2 and how the extent of discrimination relates to the demographic characteristics of the employer in Section 3.3. In these latter sections, we focus on the decisions from the *Hiring Stage* where employers can choose between a female-even-month worker and a male-odd-month worker with identical easy round performances. We focus on these decisions for two reasons. First, since these decisions occur after the *Information Stage*, all group-level information has been provided and beliefs have been elicited. Second, if female-even-month workers are hired less than 50% of the time in these decisions, it is clear that they are discriminated against because of their association with the lower performing group.[8]

Lastly, we explore whether our results extend to the remaining decisions that employers make during our study. In particular, Section 3.4 details the hiring decisions that employers make during the *Information Stage* as well as the hiring decisions that employers make when the female-even-month worker has a better easy round performance than the male-odd-month worker in the *Hiring Stage*.

3.1 What beliefs do employers hold?

Figure 2 displays the kernel densities of employers' prior beliefs about the performance gap on the hard and easy quizzes between male-odd-month workers and female-even-month workers. As we expect, employers initially believe that male workers outperform female workers by a significantly larger amount than odd-month workers outperform even-month workers (see Panels

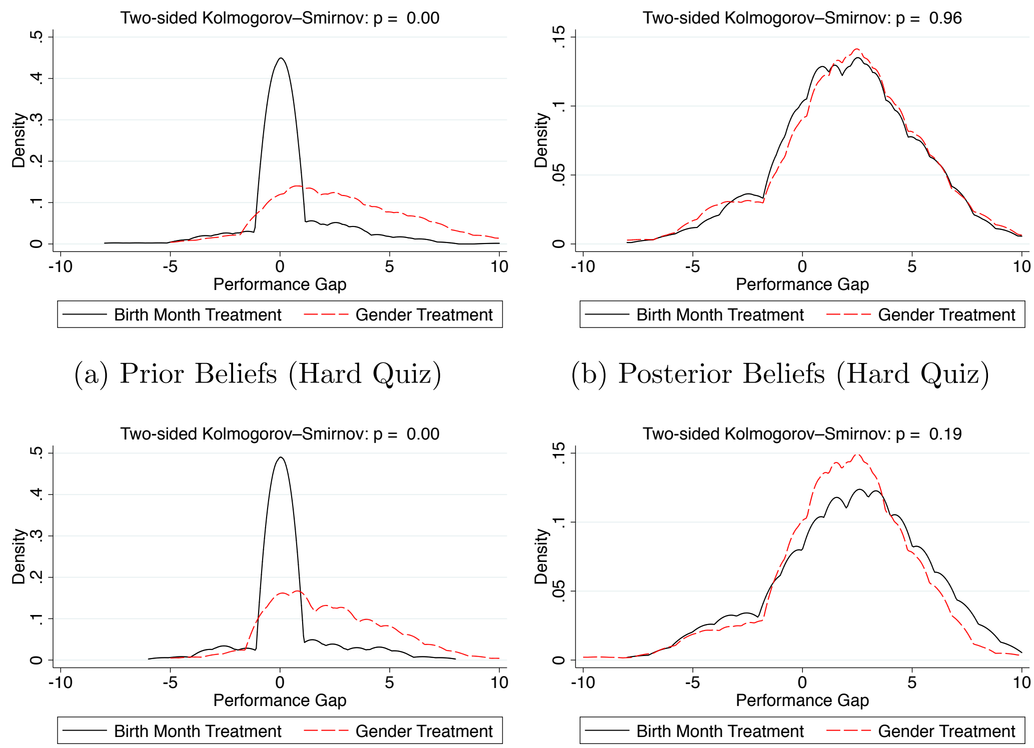
[7] Danilov and Saccardo (2017) use a related approach in their paper on ethnic discrimination. While they find no evidence of discrimination by Germans toward Turks in a basic dictator game, when they require that the sender tell a lie in order to provide the receiver with the generous outcome, Germans are less likely to do this for a Turk than a German.

[8] If the female-even-month worker is hired 50% of the time, our results would be consistent with (i) no discrimination, or (ii) a netting out different types of discrimination.

A and C). After employers complete the *Information Stage*, however, employers' posterior beliefs about the performance gap on neither the hard nor easy quiz are significantly different across treatments (see Panel B and D).[9]

These posterior beliefs make clear the motive for belief-based discrimination: employers in both treatments believe male-odd-month workers outperform female-even-month workers, on average. These posterior beliefs are also directionally accurate: male-odd-month workers outperform female-even-month workers on the both the hard and easy quizzes.[10] In fact, even when controlling for each easy quiz performance level, male-odd-month workers outperform female-even-month workers on the hard quiz.[11]

Figure 2: Beliefs about performance gap between male-odd-month workers and female-even-months workers



(a) Prior Beliefs (Hard Quiz) (b) Posterior Beliefs (Hard Quiz)

(c) Prior Beliefs (Easy Quiz) (d) Posterior Beliefs (Easy Quiz)

[9] Appendix Table A.2 confirms that this belief convergence is on average statistically significant and shows that it is is similar when separately considering workers' performance on sports or math

[10] In the sports version, relative to female-even-month workers, male-odd-month workers answer an average of 1.1 more questions correctly on the hard quiz ($p = 0.052$) and an average of 1.0 more questions correctly on the easy quiz ($p = 0.126$). In the math version, relative to female-even-month workers, male-odd-month workers answer an average of 0.451 more questions correctly on the hard quiz ($p = 0.089$) and an average of 0.692 more questions correctly on the easy quiz ($p = 0.242$).

[11] In particular, a regression of the hard round performance on an indicator for female-even-month workers and indicators for each easy round performance level indicates that male-even-month workers, relative to female-even-month workers, answer an average of 0.422 more questions correctly on the hard sports quiz ($p = 0.464$) and an average of 0.538 more questions correctly on the hard math quiz ($p = 0.089$).

3.2 Is there discrimination against women?

We begin by considering hiring decisions between a male-odd-month worker and a female-even-month worker when all else is equal. Employers learn that the payoff rule from hiring either worker is the same: they receive 10 cents per correct answer on the hard quiz. Employers also receive identical performance information on both workers: both workers are known to answer 4, 6, or 8 (out of 10) questions correctly on the easy quiz.

If employers neither engage in belief-based nor taste-based discrimination, we expect that they should hire female-even-month workers 50% of the time.[12] This is not the case. Employers in the Gender treatment only hire female workers in 43% of decisions, significantly below the 50% benchmark.

This discrimination is not specific to gender, however. In the Birth Month treatment, an even-month worker is hired in just 37% of decisions, also significantly below the 50% benchmark. In other words, employers in both treatments discriminate against workers associated with the lower-performing group, consistent with beliefs driving discrimination against women, rather than tastes.

Indeed, Column 1 of Table 1 Panel A confirms that taste-based considerations, if anything, work in favor of women: female-even-month workers are 6 percentage points more likely to be hired when they are labeled as female workers in the Gender treatment than when they are labeled as even-month workers in the Birth Month treatment.

In Column 2 of Table 1 Panel A, we directly explore the role of beliefs. We add to the regression two belief measures: the employer's posterior belief of the average performance gap in the easy quiz, and her imputed belief of "differential improvement" from the easy to the hard quiz.[13] Both measures are signed so that a positive gap indicates a belief of a male advantage. Column 2 shows that beliefs are highly predictive of decisions: employers who believe the performance gap is larger are less likely to hire the worker from the lower-performing group.[14]

Columns 3 - 4 of Table 1 Panel A present results from when a "veil" for employers intentions is provided. We do this via the addition of risk into the payoff from hiring a female-even-month worker but not a male-odd-month worker. Two subsequent findings are of note. First, women labeled as female rather than even-month are no longer significantly more likely to be hired.

[12] We code the probability of a female-even-month workers being hired as 1 if she is hired with certainty, 0.5 if chance determines who is hired, and 0 if a male-odd-month worker is instead hired with certainty.

[13] Differential improvement is the idea that participants may believe that the easy quiz performance is differentially predictive of the hard quiz performance across the two groups. To measure this, we difference the participant's belief of the hard quiz performance gap and the easy quiz performance gap, and include this difference-in-differences as a control. By including both differential improvement and posterior belief of the easy quiz gap, we allow our specifications to depend on beliefs about both the easy and hard quiz. Our results are not sensitive to this particular construction. Results available upon request. Of course, these belief measures are unincentivized, and we lack data on the confidence with which employers hold them.

[14] Column 1 of Appendix Table A.3 shows that beliefs are not more predictive in the Gender treatment, even though beliefs about gender differences could be more strongly held than beliefs about birth month differences.

Table 1: Hiring decisions between workers with the same performance

	Same payoff rule		Risk in payoff rule only for female-even-month workers	
	(1)	(2)	(3)	(4)
Panel A: Gender and Birth Month Treatments				
Gender Treatment	0.061***	0.057***	0.026	0.024
	(0.019)	(0.018)	(0.019)	(0.018)
Posterior(easy gap)		-0.023***		-0.018***
		(0.003)		(0.003)
Posterior(hard-easy gap)		-0.011**		-0.011**
		(0.005)		(0.004)
Decision FEs	yes	yes	yes	yes
Observations	2400	2400	12000	12000
Panel B: Gender and Gender-No-Information Treatments				
No Information	0.002	0.004	0.005	0.008
	(0.020)	(0.019)	(0.019)	(0.019)
Posterior(easy gap)		-0.021***		-0.015***
		(0.004)		(0.004)
Posterior(hard-easy gap)		-0.014***		-0.015***
		(0.005)		(0.004)
Decision FEs	yes	yes	yes	yes
Observations	2409	2409	12045	12045

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making hiring decisions between pairs of workers who have the same performances. ***Gender Treatment*** is an indicator for employers in the Gender treatment. ***Posterior(easy gap)*** is an employer's posterior belief about the average performance gap in the easy quiz. ***Posterior(hard − easy gap)*** is an employer's posterior belief about the average performance gap in the hard quiz minus the employer's posterior belief about the average performance gap in the easy quiz. ***No Information*** is an indicator for employers in the No Information Stage version. Decisions FEs include 6 fixed effects (with one excluded) that capture whether hiring decisions are between workers with equal performances on the sports quiz (4 v 4, 6 v 6, or 8 v 8) or between workers with equal performances on the math quiz (4 v 4, 6 v 6, or 8 v 8). Columns 1 - 2 involve hiring decisions where the payoff rule is always the same, while Columns 3 - 4 involve hiring decisions where the payoff rule is risky when hiring female-even-month workers and riskless when hiring male-odd-month workers. Panel A involves hiring decisions from the Birth Month and Gender treatments, while Panel B involves hiring decisions from the Gender and Gender-No-Information treatments.

Thus, the results under risk suggest that some of the female preference we observe in the stark, riskless environment may be driven by image concerns.[15] Indeed, the odd columns in Appendix Table A.5 show we do not observe evidence for the female preference when separately examining our results for each level of risk we introduce into decisions, including when risk only entails a 1%

[15] While the preference for hiring women relative to even-month is insignificant under risk, we cannot reject that the effects are the same in the risk and riskless decisions. See Columns 1 and 2 of Appendix Table A.4.

chance of no payment from hiring the female-even-month worker. Second, and more importantly for our purposes, female-even-month workers are still *not less likely* to be hired under risk. Thus, gender-based animus against women does not arise even when a veil for employers' intentions exists. It does not appear to be the case that the absence of taste-based discrimination against women in riskless decisions was only an artifact of image concerns or social desirability bias.

Taken together, our results may be summarized as follows. When women are the lower-performing group, they are discriminated against even when their individual performances are as good as their counterparts. This type of discrimination, however, is not specific to their gender. Rather, it appears driven by beliefs. Female workers are treated no worse than even-month workers, even when there is a veil for employers' intentions.

After seeing these results, we were curious about the extent to which the observed discrimination against women was driven by the fact that participants receive ample, detailed information about the average gender gap in performance in the easy quiz. We hypothesized that this information may have helped participants feel justified in their decision to discriminate. Would participants be willing to discriminate without this information?

To test this hypothesis, we ran a new "No Information Stage" version of the experiment that simply removed the Information Stage for employers. All other aspects of the design were kept the same, and we again recruited 800 individuals from Amazon Mechanical Turk. These results are presented in Panel B of Table 1, both without risk (Columns 1 - 2) and with risk (Columns 3 - 4). We ask whether female workers are hired as often in the Gender treatment of the No Information Stage version as in our baseline Gender treatment. Column 1 shows that the level of discrimination against women is identical without the Information Stage; this remains true when we control for beliefs in Column 2, and when we consider the risky decisions in Columns 3 and 4. Providing ample information about differences by gender neither enables nor promotes discrimination against women above and beyond what we find when employers are acting only on their priors.[16] Employers are willing to discriminate even based on arguably less informed beliefs.

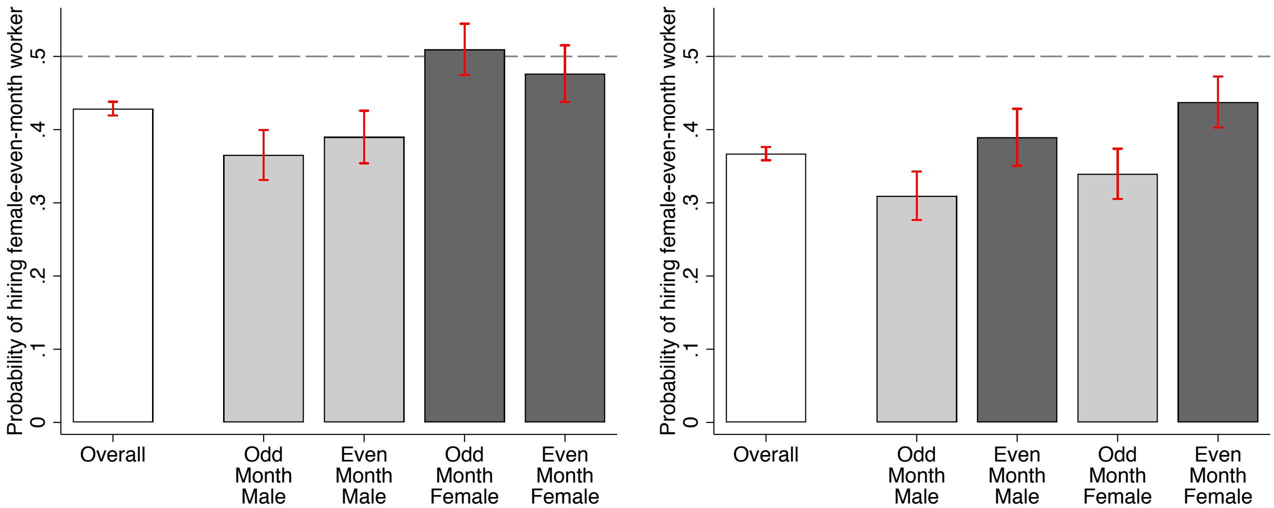
3.3 Does the gender of the employer matter?

Figure 3 shows that there is a substantial heterogeneity in discrimination when considering the four types of employers in our experiment as defined by their gender and their birth month. In the Gender treatment, male employers engage in substantial discrimination against female workers, hiring them less than 40% of the time. By contrast, and consistent instead with no discrimination, female employers hire female workers approximately 50% of the time.[17]

[16] Not surprisingly, discrimination against even-month workers no longer persists absent the Information stage (see Appendix Figure A.1).

[17] Past evidence on whether female employers are in general more likely to hire female workers is mixed, see for instance Bagues and Esteve-Volart (2010) and De Paola and Scoppa (2015).

Figure 3: Hiring decisions between workers with same performance and payoff rule



(a) Gender Treatment

(b) Birth Month Treatment

Just as discrimination against female workers is not specific to gender, however, this heterogeneity in discrimination is not specific to gender. While even-month employers hire odd-month workers approximately 30% of the time, this discrimination is less (but persistent) when considering the hiring decisions of even-month employers, who hire even-month workers approximately 40% of the time. In other words, it is not simply that female employers discriminate less often against female workers. Employers discriminate less often against "in-group" members more broadly. This is consistent with previous work on in-group preference that documents that shared social identities can impact economic outcomes, including studies using both minimal group paradigms and natural identities (see Tajfel et al. (1971), Brewer (1979), Chen and Li (2009), Chen and Chen (2011), and Chen et al. (2014)). We explore this pattern in Table 2. Column 1 documents a statistically significant in-group preference, while Column 2 shows that the extent of in-group preference does not vary across treatment. In Column 3, we show that this preference is not explained by employer beliefs.

However, in our setting, evidence for an in-group preference is not robust. When provided with a veil via risk on their intentions, employers do not demonstrate an in-group preference (see Columns 4 - 6 of Table 2, and the odd columns in Appendix Table A.5 to examine the results separately for each level of risk).[18] This suggests that the in-group preference in the stark environment of the riskless decisions is likely reflective of image concerns, including potentially experimenter demand.

[18]See Columns 3 and 4 of Appendix Table A.4 for an interacted model.

Table 2: Hiring decisions between workers with the same performance

	Same payoff rule			Risk in payoff rule only for female-even-month workers		
	(1)	(2)	(3)	(4)	(5)	(6)
In-Group	0.098***	0.090***	0.081***	0.012	0.008	0.001
	(0.019)	(0.026)	(0.025)	(0.019)	(0.025)	(0.024)
Gender Treatment		0.055**	0.057**		0.022	0.025
		(0.025)	(0.025)		(0.025)	(0.025)
*Gender Treatment*In-Group*		0.020	0.007		0.009	-0.002
		(0.037)	(0.036)		(0.038)	(0.037)
Posterior(easy gap)			-0.021***			-0.018***
			(0.003)			(0.003)
Posterior(hard-easy gap)			-0.011**			-0.011**
			(0.005)			(0.004)
Decision FEs	yes	yes	yes	yes	yes	yes
Observations	2400	2400	2400	12000	12000	12000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making hiring decisions between pairs of workers who have the same performances. ***In-Group*** is an indicator for employers who know they share some demographic characteristic with the female-even-month workers (i.e., even-month employers in the Birth Month treatment and female employers in the Gender treatment). ***Gender Treatment*** is an indicator for employers in the Gender treatment. ***Posterior(easy gap)*** is an employer's posterior belief about the average performance gap in the easy quiz. ***Posterior(hard − easy gap)*** is an employer's posterior belief about the average performance gap in the hard quiz minus the employer's posterior belief about the average performance gap in the easy quiz. Decisions FEs include 6 fixed effects (with one excluded) that capture whether hiring decisions are between workers with equal performances on the sports quiz (4 v 4, 6 v 6, or 8 v 8) or between workers with equal performances on the math quiz (4 v 4, 6 v 6, or 8 v 8). Columns 1 - 3 involve hiring decisions where the payoff rule is always the same, while Columns 4 - 6 involve hiring decisions where the payoff rule is risky when hiring female-even-month workers and riskless when hiring male-odd-month workers. All hiring decisions are from the Birth Month and Gender treatments.

3.4 Do our results extend to other decisions?

In Sections 3.2 and 3.3, we focus on hiring decisions between two workers with identical easy round performances. It is in these decisions that we see clear evidence of discrimination: differential treatment by employers of two employees with identical performance information. In this section, we examine whether our results extend to additional hiring decisions where: (i) the individual performances of the two workers are unknown, or (ii) the female-even-month worker has a better easy round performance than the male-odd-month does.

The first set of additional hiring decisions involves the 12 decisions that employers make as part of the *Information Stage*. In each of these decisions, the distribution of easy round performances for some subset of female-even-month workers and the distribution of easy round performances for some subset of male-odd-month workers are displayed. Employers then choose between a female-even-month worker and male-odd-month worker, knowing that their worker will be randomly drawn from the subset of workers comprising the relevant, displayed distribution. As shown in Appendix Table A.1, these displayed distributions, on average, largely favor male-odd-month workers. Out of the 24 distributions that are displayed in the math version or in sports version, there are only 3 distributions that favor the female-even-month workers and 1 distribution that neither favors the male-odd-month workers nor the female-even-month workers in terms of average performance difference.

Here, we focus on the decisions over the 20 distributions for which the average performance gap favors the male-odd-month workers.[19] The corresponding decisions can be summarized as follows (and see Appendix Figure A.2 and Appendix Table A.6 for reference). First, employers are less likely to hire from the lower-performing distribution: female workers in the Gender treatment are hired 33% of the time when the average performance gap of the displayed distributions favors male workers. Second, this difference is not specific gender: even-month workers in the Birth Month treatment are hired 29% of the time, significantly lower than the 33% of the time that female workers are hired, when the average performance gap of the displayed distributions favors odd-month workers. Third, as in our riskless decisions over workers with identical performances, evidence for an in-group preference is substantial and significant. Female employers are 16 percentage points more likely to hire female workers in the Gender treatment, and even-month employers are 6 percentage points more likely to hire even-month workers in the Birth Month treatment. Indeed, the slight preference for hiring female, as compared to even-month, workers appears entirely driven by a stronger in-group preference.

The second set of additional hiring decisions involve employers in the *Hiring Stage* choosing between pairs of workers in which the female-even-month worker has a better easy round performance than the male-odd-month worker. The corresponding decisions can be summarized as follows (and see Appendix Figure A.5 as well as Appendix Table A.9 for reference). First, female-even-month workers are hired nearly 90% of the time in both of our treatments, suggesting that any discrimination against female-even-month workers is not so strong as to induce employers not to hire a worker with the better easy round performance. Second, we again do not observe any evidence consistent with taste-based animus against female workers: female workers in the Gender treatment are not less likely to be hired than even-month workers in the Birth Month treatment. Third, we do not observe any significant differences according to the demographics

[19] As one would expect, the results are much noisier when only considering the 3 distributions for which the average performance gap favors the female-even-month worker (see Appendix Figure A.3 and Appendix Table A.7) and the 1 distribution for which there is no the average performance gap (see Appendix Figure A.4 and Appendix Table A.8).

of the employer, suggesting that any in-group preference is not so strong as to overcome a desire to hire the worker with the better easy round performance.

4 Conclusion

We leverage a controlled environment in order to disentangle the motivations behind discrimination against women in a male-typed employment environment. We find ample evidence of discrimination against women, as employers are significantly less likely to hire a woman compared to an equally able man.

This discrimination, however, is not specific to gender. When workers are identified as members of one of two groups, we observe similar levels of discrimination against the lower-performing group regardless of whether they are identified according to gender or birth month. This evidence points to an important role for beliefs about performance in explaining discriminatory practices. Theories of stereotyping and belief formation, like those of Bordalo et al. (2016), may be particularly important for understanding and reducing discrimination.

We also provide evidence on in-group preferences, finding that male employers, more so than female employers, discriminate against female workers in the Gender treatment. Again however, this finding is not specific to gender. Odd-month employers, more so than even-month employers, discriminate against even-month workers. This evidence highlights how discrimination can be mitigated by an in-group preference or activated by an out-group animus. We caution, however, that when employers are provided with a veil on their intentions, the addition of risk to the decision, we no longer observe a significant role for in-group preferences.

Put simply, our paper documents the value of careful control comparisons to narrow in on the drivers of discrimination. Providing ample information on the distribution of performances for a group — but varying whether the gender of that group is known — is a powerful way to examine the role of beliefs related to average group differences. Average group differences are central to many theories of discrimination; collecting data that speaks directly to their importance provides important new evidence around these theories. Labeling a control group by an arbitrary characteristic — rather than simply removing a gender label — allows us to narrow in on gender-specific channels. It prevents us from attributing taste-based considerations that are driven by in-group preferences to gender. Introducing a veil for intentions — given that hiring decisions outside of the laboratory frequently involve such noise or plausible veils — speaks to the robustness of findings. Neither female, as compared to even-month, workers nor in-group, as compared to out-group, members are more likely to be hired once risk provides a plausible veil for employers' intentions. Thus, potential veils for intentions, such as risk, may be a useful methodological tool when seeking to minimize the role of image concerns, including those related to social desirability or experimenter demand concerns, in driving results.

We conclude with four additional points. First, our findings show that beliefs and tastes may work in opposite directions in driving discrimination. When considering the overall results in our context, beliefs push employers to hire women less often, while tastes, if anything, seem to push employers to hire women more often. When considering the heterogeneity in our results, different taste-based considerations clearly push in opposite directions.

Second, our findings underscore the challenges faced by members of a group that is believed to be lower-performing on average. Mere membership in a lower-performing group — even when this membership is outside of the control of the worker and based on an arbitrary characteristic — is sufficient for discrimination to follow. This challenge, moreover, is clearly exacerbated for members of a group that is not well-represented among decision-makers. For instance, had all employers been males in our study, the overall level of discrimination against female workers would have been greater.

Third, despite the potential reluctance to appear sexist, evidence for discrimination against women occurs even in a stark environment. In our study, employers make a side-by-side decision between a male worker and a female worker with identical resumes, there is little ambiguity to use to justify discrimination, and it is quite cheap to not appear sexist. And, yet, we find that employers, on average, are willing to discriminate against women.

Fourth, while using birth month to label groups may be an unusual approach, it serves our purposes well. And while association with birth month is not a commonly used feature outside of the laboratory, we note that other group associations based upon school affiliation, section or classroom assignment, workplace group, office location, or team are quite common. Our paradigm suggests that these associations may play a non-trivial role in how students and workers from these groups are evaluated and thus the extent to which they may serve as appropriate controls when seeking to narrow in on the drivers of discrimination.

References

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Chapter 48 Race and gender in the labor market." In . Vol. 3 of *Handbook of Labor Economics*, 3143 – 3259. Elsevier.

Arrow, Kenneth. 1973. "Discrimination in Labor Markets." , ed. Orley Ashenfelter and Albert Rees, Chapter The Theory of Discrimination. Princeton, NJ:Princeton University Press.

Ayres, Ian, and Peter Siegelman. 1995. "Race and gender discrimination in bargaining for a new car." *The American Economic Review*, 304–321.

Azmat, Ghazala, and Barbara Petrongolo. 2014. "Gender and the labor market: What have we learned from field and lab experiments?" *Labour Economics*, 30: 32–40.

Baert, Stijn, Ann-Sophie De Pauw, and Nick Deschacht. 2016. "Do employer preferences contribute to sticky floors?" *ILR Review*, 69(3): 714–736.

Bagues, Manuel F, and Berta Esteve-Volart. 2010. "Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment." *The Review of Economic Studies*, 77(4): 1301–1328.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." *American Economic Review*, 106(6): 1437–1475.

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

Bertrand, Marianne. 2011. "New perspectives on Gender." *Handbook of Labor Economics*, 4: 1543–1590.

Bertrand, Marianne, and Esther Duflo. 2017. "Field experiments on discrimination." *Handbook of Economic Field Experiments*, 1: 309–393.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *The American Economic Review*, 94(4): 991–1013.

Black, Sandra E, and Philip E Strahan. 2001. "The division of spoils: rent-sharing and discrimination in a regulated industry." *The American Economic Review*, 91(4): 814–831.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of Economic Literature*, 3: 789–865.

Bohnet, Iris, Alexandra van Geen, and Max Bazerman. 2016. "When Performance Trumps Gender Bias: Joint Versus Separate Evaluation." *Management Science*, 62(5): 1225–1234.

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg. 2017. "The Dynamics of Discrimination: Theory and Evidence." *Working Paper.*

Bordalo, Pedro, Katherine B. Coffman Nicola Gennaioli, and Andrei Shleifer. 2018. "Beliefs about Gender." *Working Paper.*

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *The Quarterly Journal of Economics*, 131(4): 1753–1794.

Bowles, Hannah Riley, Linda Babcock, and Lei Lai. 2007. "Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask." *Organizational Behavior and Human Decision Processes*, 103(1): 84–103.

Brewer, Marilynn B. 1979. "In-group bias in the minimal intergroup situation: A cognitive-motivational analysis." *Psychological bulletin*, 86(2).

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women." *Quarterly Journal of Economics*, 131(2): 633–686.

Castillo, Marco, and Ragan Petrie. 2010. "Discrimination in the lab: Does information trump appearance?" *Games and Economic Behavior*, 68.

Castillo, Marco, Ragan Petrie, Maximo Torero, and Lise Vesterlund. 2013. "Gender differences in bargaining outcomes: A field experiment on discrimination." *Journal of Public Economics*, 99: 35–48.

Charles, Kerwin Kofi, and Jonathan Guryan. 2008. "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination." *Journal of Political Economy*, 116(5): 773–809.

Charles, Kerwin Kofi, and Jonathan Guryan. 2011. "Studying discrimination: Fundamental challenges and recent progress." *Annual Review of Economics*, 3(1): 479–511.

Chen, Roy, and Yan Chen. 2011. "The potential of social identity for equilibrium selection." *The American Economic Review*, 101(6): 2562–2589.

Chen, Yan, and Sherry Xin Li. 2009. "Group identity and social preferences." *The American Economic Review*, 99(1): 431–457.

Chen, Yan, Sherry Xin Li, Tracy Xiao Liu, and Margaret Shih. 2014. "Which hat to wear? Impact of natural identities on coordination and cooperation." *Games and Economic Behavior*, 84: 58–86.

Coffman, Katherine Baldiga. 2014. "Evidence on Self-Stereotyping and the Contribution of Ideas." *The Quarterly Journal of Economics*, 129(4): 1625–1660.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature*, 47(2): 448–474.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang. 2007. "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness." *Economic Theory*, 33: 67–80.

Danilov, Anastasia, and Silvia Saccardo. 2017. "Disguised Discrimination." *Working Paper.*

De Paola, Maria, and Vincenzo Scoppa. 2015. "Gender discrimination and evaluators' gender: evidence from Italian academia." *Economica*, 82(325).

Dickinson, David L., and Ronald L. Oaxaca. 2009. "Statistical Discrimination in Labor Markets: An Experimental Analysis." *Southern Economic Journal*, 76(1): 16–31.

Dickinson, David L, and Ronald L. Oaxaca. 2014. "Wages, Employment, and Statistical Discrimination Evidence from the laboratory." *Economic Inquiry*, 52(4): 1380–1391.

Doleac, Jennifer L., and Luke C.D. Stein. 2013. "The Visible Hand: Race and Online Market Outcomes." *The Economic Journal*, 123.

Exley, Christine L. 2015. "Excusing Selfishness in Charitable Giving: The Role of Risk." *Review of Economic Studies*, 83(2): 587–628.

Fershtman, Chaim, and Uri Gneezy. 2001. "Discrimination in a segmented society: An experimental approach." *The Quarterly Journal of Economics*, 116(1): 351–377.

Gino, Francesca, Michael I. Norton, and Roberto A. Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives*, 30(3): 189–212.

Gneezy, Uri John List, Michael K. Price. 2012. "Toward an Understanding of Why People Discriminate: Evidence from a Series of Natural Field Experiments." *NBER Working Paper.*

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1091–1119.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of" Blind" Auditions on Female Musicians." *The American Economic Review*, 90(4): 715–741.

Guryan, Jonathan, and Kerwin Kofi Charles. 2013. "Taste-based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots." *The Economic Journal*, 123(572).

Haisley, Emily C., and Roberto A. Weber. 2010. "Self-serving interpretations of ambiguity in other-regarding behavior." *Games and Economic Behavior*, 68: 614–625.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin*, 108(3): 480–498.

Lei, Lai, and Linda C. Babcock. 2013. "Asian Americans and workplace discrimination: The interplay between sex of evaluators and the perception of social skills." *Journal of Organizational Behavior*, 34(3): 310–326.

List, John A. 2004. "The nature and extent of discrimination in the marketplace: Evidence from the field." *The Quarterly Journal of Economics*, 119(1): 48–89.

Milkman, Katherine L., Modupe Akinola, and Dolly Chugh. 2012. "Temporal Distance and Discrimination: An Audit Study in Academia." *Psychological Science*, 23(7): 710–717.

Milkman, Katherine L, Modupe Akinola, and Dolly Chugh. 2015. "What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations." *Journal of Applied Psychology*, 100(6).

Mobius, Markus M, and Tanya S Rosenblat. 2006. "Why beauty matters." *American Economic Review*, 96(1): 222–235.

Moss-Racusin, Corinne A, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman. 2012. "Science faculty's subtle gender biases favor male students." *Proceedings of the National Academy of Sciences*, 109(41): 16474–16479.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." *Quarterly Journal of Economics*, 111(3): 915–941.

Niederle, Muriel. 2016. "Gender." In *Handbook of Experimental Economics*. Vol. 2, , ed. John Kagel and Alvin E. Roth, 481–553. Princeton University Press.

Olivetti, Claudia, and Barbara Petrongolo. 2016. "The Evolution of Gender Gaps in Industrialized Countries." *Annual Review of Economics*, 8(1): 405–434.

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

Rao, Gautam. 2018. "Familiarity Does Not Breed Contempt: Generosity, Discrimination and Diversity in Delhi." *Working Paper.*

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How stereotypes impair women's careers in science." *Proceedings of the National Academy of Sciences*, 111(12): 4403–4408.

Riach, P. A., and J. Rich. 2002. "Field Experiments of Discrimination in the Market Place." *The Economic Journal*, 112(483).

Sarsons, Heather. 2017*a*. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals." *Working Paper.*

Sarsons, Heather. 2017*b*. "Recognition for Group Work: Gender Differences in Academia†." *American Economic Review: Papers & Proceedings*, 107(5): 141–145.

Tajfel, Henri, Michael G Billig, Robert P Bundy, and Claude Flament. 1971. "Social categorization and intergroup behaviour." *European Journal of Social Psychology*, 1(2): 149–178.

Appendices

A Additional Results

Table A.1: Descriptive statistics on performance distributions seen in Part 1

	Performance distribution in easy round of math quiz											
	1	2	3	4	5	6	7	8	9	10	11	12
Avg Male Advantage	2.10	2.07	1.62	1.27	0.91	0.65	0.50	0.00	-0.38	-0.43	-0.80	0.69
Men FOSD Women	yes	yes	yes	yes	no	no	no	no	no	no	no	no
Workers born on days	1-5	21-25	21-31	26-31	16-31	1-10	1-15	11-15	11-20	16-20	6-10	1-31
Number of workers	9	8	19	11	30	16	20	4	15	11	7	50
	Performance distribution in easy round of sports quiz											
	1	2	3	4	5	6	7	8	9	10	11	12
Avg Male Advantage	4.08	2.50	1.34	2.25	2.08	1.85	0.73	0.69	0.72	0.50	0.24	1.00
Men FOSD Women	yes	yes	yes	no	no	no	no	no	no	no	no	no
Workers born on days	21-31	11-15	16-20	21-31	6-10	16-31	1-10	1-15	11-20	21-27	1-5	1-31
Number of workers	7	7	6	10	7	16	34	27	13	5	20	50

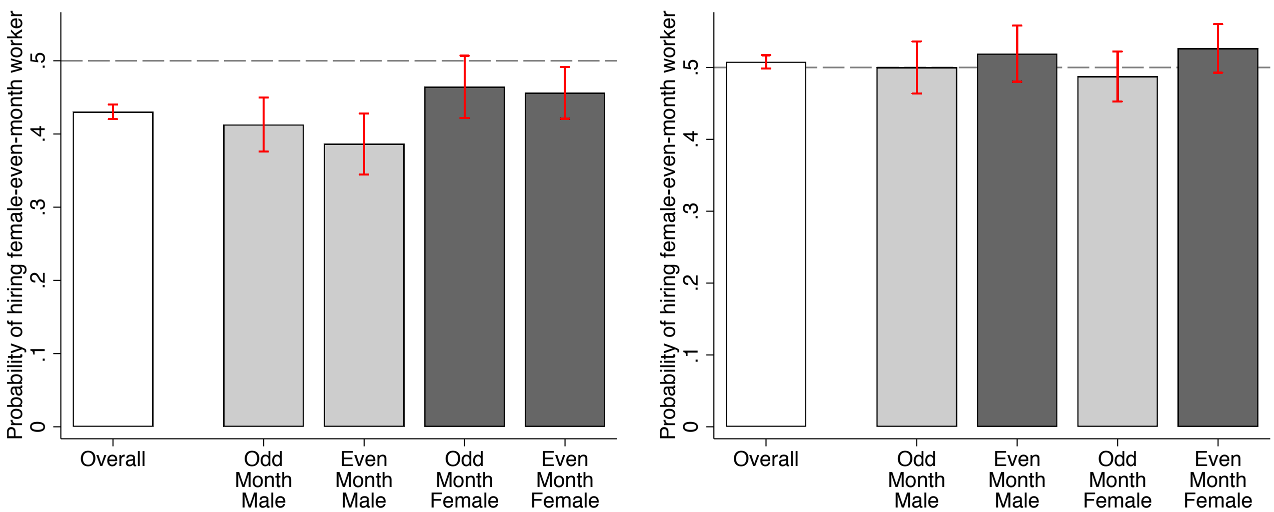
Avg Male Advantage is the average performance of male workers minus the average performance of female workers. **Men FOSD Women** indicates whether the performance of male workers first order stochastically dominates the performance of female workers. **Workers born on days** indicates the days during which workers in a particular distribution were born. **Number of workers** indicates how many workers are involved in a particular distribution (note that employers are not aware of the sample size for each distribution however).

Table A.2: Beliefs about the performance gap between male-odd-months workers and female-even-month workers

	Hard quiz performance gap			Easy quiz performance gap		
	Prior	Posterior	Prior − Posterior	Prior	Posterior	Prior − Posterior
	(1)	(2)	(3)	(4)	(5)	(6)
Math and Sports Version						
Gender Treatment	2.344***	0.031	-2.313***	1.845***	-0.368	-2.213***
	(0.177)	(0.220)	(0.256)	(0.159)	(0.225)	(0.243)
Constant	0.387***	2.121***	1.734***	0.327***	2.209***	1.882***
	(0.095)	(0.153)	(0.183)	(0.090)	(0.167)	(0.181)
Observations	800	800	800	800	800	800
Math Version						
Gender Treatment	0.997***	-0.354	-1.352***	0.597***	-0.906***	-1.502***
	(0.229)	(0.294)	(0.358)	(0.194)	(0.296)	(0.316)
Constant	0.389***	1.899***	1.510***	0.374***	1.980***	1.606***
	(0.136)	(0.217)	(0.273)	(0.124)	(0.222)	(0.243)
Observations	400	400	400	400	400	400
Sports Version						
Gender Treatment	3.705***	0.425	-3.280***	3.105***	0.180	-2.925***
	(0.234)	(0.322)	(0.359)	(0.222)	(0.331)	(0.365)
Constant	0.385***	2.340***	1.955***	0.280**	2.435***	2.155***
	(0.134)	(0.215)	(0.245)	(0.131)	(0.248)	(0.268)
Observations	400	400	400	400	400	400

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from an ordinary least squares model of the beliefs about the extent of the performance gap between male-odd-month workers and female-even-month workers. ***Gender Treatment*** is an indicator for when workers are labeled according to their gender as opposed to birth month. Results are from the Birth Month and Gender treatments.

Figure A.1: In the No Information Stage version, hiring decisions between workers with same performance and payoff rule



(a) Gender Treatment

(b) Birth Month Treatment

Table A.3: With belief interactions: hiring decisions between workers with the same performance

	Same payoff rule (1)	Risk in payoff rule only for female-even-month workers (2)
Panel A: Gender and Birth Month Treatments		
Gender Treatment	0.044**	0.028
	(0.022)	(0.024)
Posterior(easy gap)	-0.026***	-0.017***
	(0.004)	(0.004)
Posterior(hard-easy gap)	-0.014**	-0.007
	(0.007)	(0.005)
*Gender Treatment*Posterior(easy gap)*	0.006	-0.002
	(0.006)	(0.007)
*Gender Treatment*Posterior(hard-easy gap)*	0.005	-0.008
	(0.009)	(0.009)
Decision FEs	yes	yes
Observations	2400	12000
Panel B: Gender and Gender-No-Information Treatments		
No Information	0.011	-0.009
	(0.023)	(0.026)
Posterior(easy gap)	-0.020***	-0.019***
	(0.005)	(0.005)
Posterior(hard-easy gap)	-0.009	-0.015**
	(0.006)	(0.007)
*No Information*Posterior(easy gap)*	-0.002	0.009
	(0.007)	(0.007)
*No Information*Posterior(easy gap)*	-0.009	-0.000
	(0.009)	(0.008)
Decision FEs	yes	yes
Observations	2409	12045

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making hiring decisions between pairs of workers who have the same performance and payoff rules. ***Gender Treatment*** is an indicator for employers in the Gender treatment. ***Posterior(easy gap)*** is an employer's posterior belief about the average performance gap in the easy quiz. ***Posterior(hard − easy gap)*** is an employer's posterior belief about the average performance gap in the hard quiz minus the employer's posterior belief about the average performance gap in the easy quiz. ***No Information*** is an indicator for employers in the No Information Stage version. Decisions FEs include 6 fixed effects (with one excluded) that capture whether hiring decisions are between workers with equal performances on the sports quiz (4 v 4, 6 v 6, or 8 v 8) or between workers with equal performances on the math quiz (4 v 4, 6 v 6, or 8 v 8). Column 1 involves hiring decisions where the payoff rule is always the same, while Column 2 involves hiring decisions where the payoff rule is risky when hiring female-even-month workers and riskless when hiring male-odd-month workers. Panel A involves hiring decisions from the Birth Month and Gender treatments, while Panel B involves hiring decisions from the Gender and Gender-No-Information treatments.

Table A.4: With risk interactions: hiring decisions between workers with the same performance

	(1)	(2)	(3)	(4)
Risky Payoff	-0.174***	-0.174***	-0.152***	-0.152***
	(0.016)	(0.016)	(0.015)	(0.015)
Gender Treatment	0.061***	0.059***		
	(0.019)	(0.018)		
*Gender Treatment*Risky Payoff*	-0.035	-0.035		
	(0.023)	(0.023)		
In-Group			0.100***	0.086***
			(0.019)	(0.018)
*In-Group*Risky Payoff*			-0.088***	-0.088***
			(0.023)	(0.023)
Posterior(easy gap)		-0.019***		-0.019***
		(0.003)		(0.003)
Posterior(hard-easy gap)		-0.011***		-0.011***
		(0.004)		(0.004)
Decision FEs	yes	yes	yes	yes
Observations	14400	14400	14400	14400

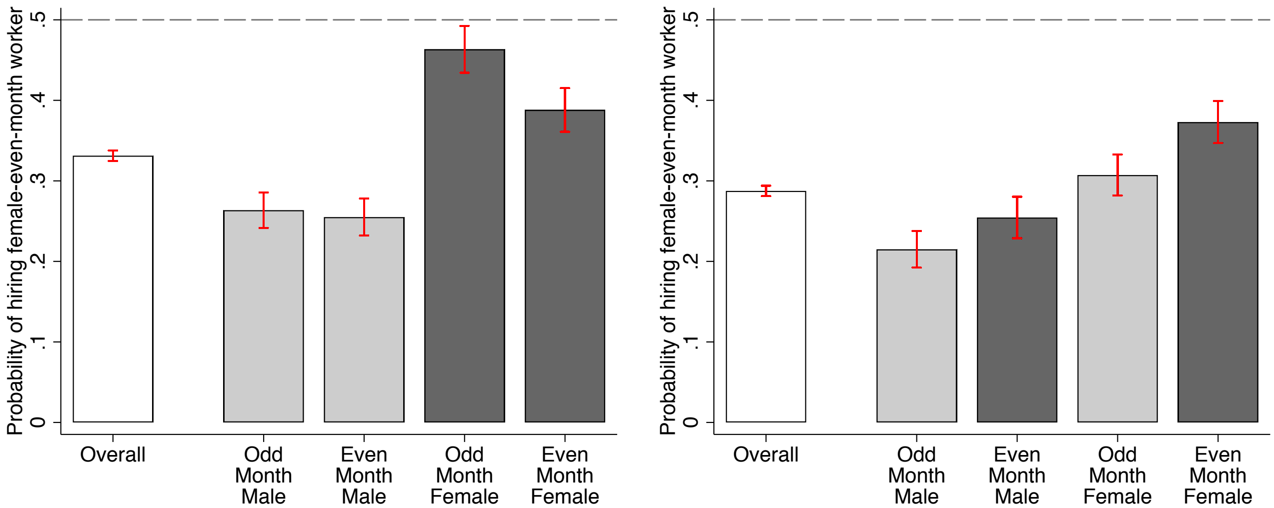
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making hiring decisions between pairs of workers who have the same performance and payoff rules. ***Risky Payoff*** is an indicator for hiring decisions where the payoff rule is risky when hiring female-even-month workers and riskless when hiring male-odd-month workers. ***Gender Treatment*** is an indicator for employers in the Gender treatment. ***In-Group*** is an indicator for employers who know they share some demographic characteristic with the female-even-month workers (i.e., even-month employers in the Birth Month treatment and female employers in the Gender treatment). ***Posterior(easy gap)*** is an employer's posterior belief about the average performance gap in the easy quiz. ***Posterior(hard − easy gap)*** is an employer's posterior belief about the average performance gap in the hard quiz minus the employer's posterior belief about the average performance gap in the easy quiz. Decisions FEs include 6 fixed effects (with one excluded) that capture whether hiring decisions are between workers with equal performances on the sports quiz (4 v 4, 6 v 6, or 8 v 8) or between workers with equal performances on the math quiz (4 v 4, 6 v 6, or 8 v 8). All hiring decisions are from the Birth Month and Gender treatments.

Table A.5: When separately considering by the risk level: hiring decisions between workers with the same performance

	Level of risk in payoff rule if a female-even-month worker is hired									
	1%		5%		10%		25%		50%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
In-Group	0.027	-0.000	0.034	0.036	0.012	0.004	-0.019	-0.018	0.006	0.019
	(0.024)	(0.033)	(0.023)	(0.032)	(0.023)	(0.031)	(0.020)	(0.027)	(0.018)	(0.024)
Gender Treatment		0.017		0.036		0.012		0.016		0.029
		(0.033)		(0.031)		(0.031)		(0.028)		(0.024)
*Gender Treatment*In-Group*		0.057		-0.000		0.017		-0.001		-0.026
		(0.048)		(0.047)		(0.046)		(0.040)		(0.036)
Decision FEs	yes	yes	yes	yes	yes	yes				
Observations	2400	2400	2400	2400	2400	2400	2400	2400	2400	2400

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making hiring decisions between pairs of workers who have the same performance but different payoff rules. ***In-Group*** is an indicator for employers who know they share some demographic characteristic with the female-even-month workers (i.e., even-month employers in the Birth Month treatment and female employers in the Gender treatment). ***Gender Treatment*** is an indicator for employers in the Gender treatment. Decisions FEs include 6 fixed effects (with one excluded) that capture whether hiring decisions are between workers with equal performances on the sports quiz (4 v 4, 6 v 6, or 8 v 8) or between workers with equal performances on the math quiz (4 v 4, 6 v 6, or 8 v 8). Columns 1 - 2, 3 -4, 5 - 6, 7 - 8, and 9 - 10, involve hiring decisions where the payoff rule involves 1%, 5%, 10%, 25%, and 50% of no payment from hiring female-even-month workers, respectively. All hiring decisions are from the Birth Month and Gender treatments.

Figure A.2: When the displayed female-even-month worker distribution has an average easy round performance that is lower: hiring decisions between workers in the *Information Stage*



(a) Gender Treatment

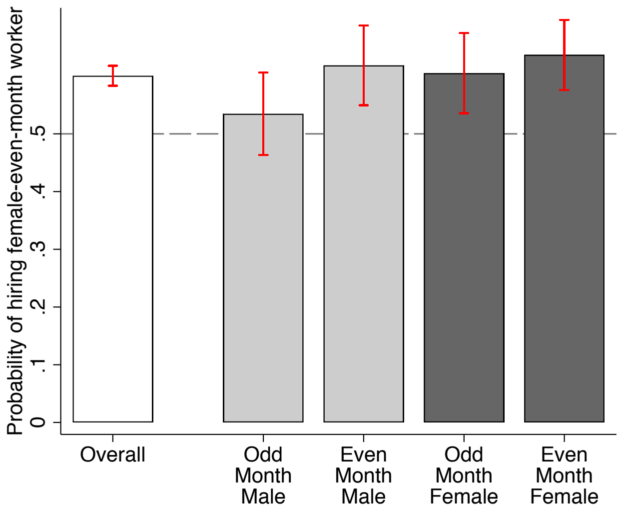
(b) Birth Month Treatment

Table A.6: When the displayed female-even-month worker distribution has an average easy round performance that is lower: hiring decisions between workers in the *Information Stage*

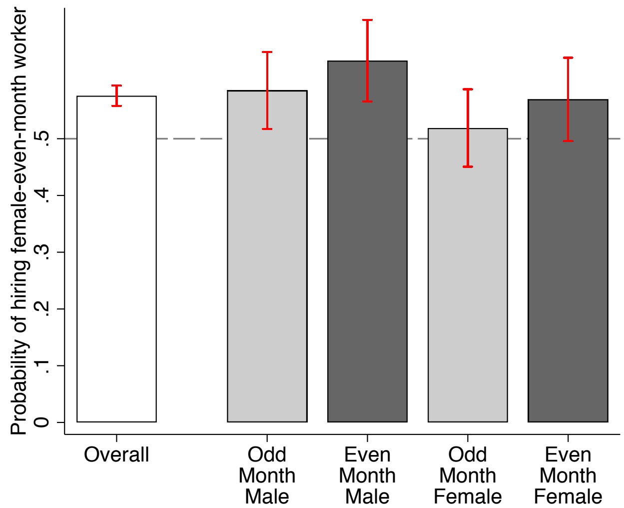
	(1)	(2)	(3)	(4)
Gender Treatment	0.044**		-0.000	0.007
	(0.019)		(0.025)	(0.021)
In-Group		0.107***	0.056**	0.032
		(0.019)	(0.027)	(0.023)
*Gender Treatment*In-Group*			0.107***	0.077**
			(0.038)	(0.032)
Posterior(easy gap)				-0.050***
				(0.003)
Posterior(hard-easy gap)				-0.024***
				(0.004)
Decision FEs	yes	yes	yes	yes
Observations	8000	8000	8000	8000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making decisions in the *Information Stage* involving displayed distributions that indicate that the average easy round performance is lower for the female-even-month workers than it is for male-odd-month workers. ***In-Group*** is an indicator for employers who know they share some demographic characteristic with the female-even-month workers (i.e., even-month employers in the Birth Month treatment and female employers in the Gender treatment). ***Gender Treatment*** is an indicator for employers in the Gender treatment. ***Posterior(easy gap)*** is an employer's posterior belief about the average performance gap in the easy quiz. ***Posterior(hard − easy gap)*** is an employer's posterior belief about the average performance gap in the hard quiz minus the employer's posterior belief about the average performance gap in the easy quiz. Decisions FEs include 20 fixed effects (with one excluded) that capture each of the 20 relevant sets of sports or math distributions. All hiring decisions are from the Birth Month and Gender treatments.

Figure A.3: When the displayed female-even-month worker distribution has an average easy round performance that is higher: hiring decisions between workers in the *Information Stage*



(a) Gender Treatment



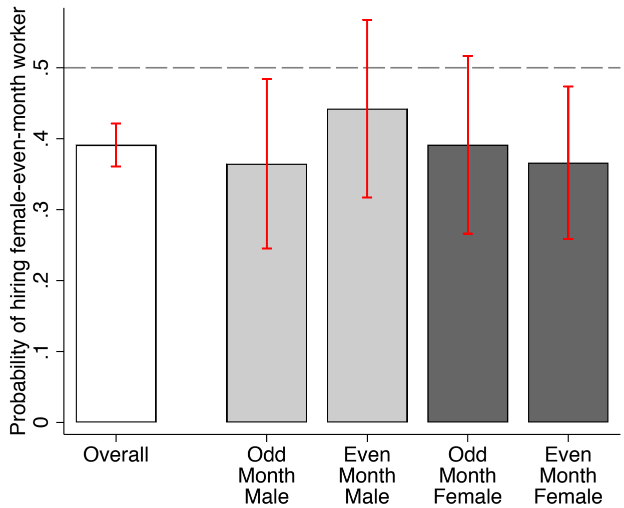
(b) Birth Month Treatment

Table A.7: When the displayed female-even-month worker distribution has an average easy round performance that is higher: hiring decisions between workers in the *Information Stage*

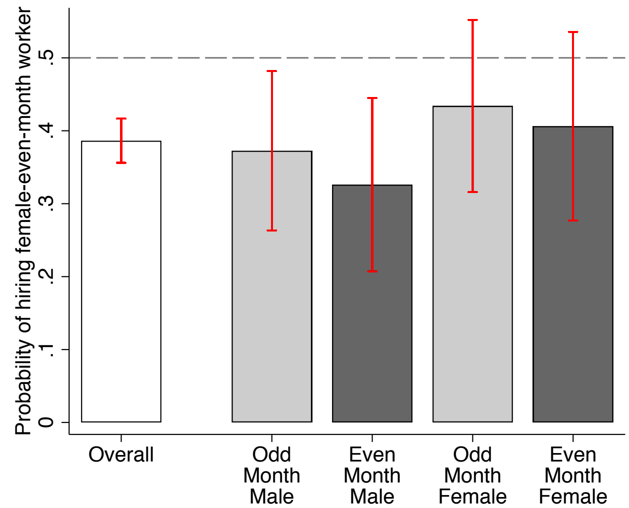
	(1)	(2)	(3)
In-Group	0.107***	0.056**	0.032
	(0.019)	(0.027)	(0.023)
In-Group	0.049**	0.052	0.050
	(0.024)	(0.035)	(0.035)
Gender Treatment		0.027	0.023
		(0.036)	(0.036)
*Gender Treatment*In-Group*		-0.007	-0.014
		(0.048)	(0.048)
Posterior(easy gap)			-0.010**
			(0.004)
Posterior(hard-easy gap)			-0.002
			(0.007)
Decision FEs	yes	yes	yes
Observations	1200	1200	1200

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making decisions in the *Information Stage* involving displayed distributions that indicate that the average easy round performance is higher for the female-even-month workers than it is for male-odd-month workers. ***In-Group*** is an indicator for employers who know they share some demographic characteristic with the female-even-month workers (i.e., even-month employers in the Birth Month treatment and female employers in the Gender treatment). ***Gender Treatment*** is an indicator for employers in the Gender treatment. ***Posterior(easy gap)*** is an employer's posterior belief about the average performance gap in the easy quiz. ***Posterior(hard − easy gap)*** is an employer's posterior belief about the average performance gap in the hard quiz minus the employer's posterior belief about the average performance gap in the easy quiz. Decisions FEs include 3 fixed effects (with one excluded) that capture each of the 3 relevant sets of sports or math distributions. All hiring decisions are from the Birth Month and Gender treatments.

Figure A.4: When there is no average difference in the average easy round performances between the displayed distributions: hiring decisions between workers in the *Information Stage*



(a) Gender Treatment



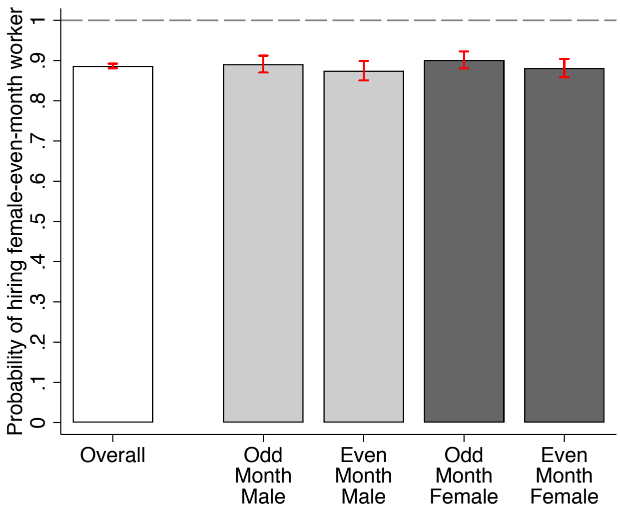
(b) Birth Month Treatment

Table A.8: When there is no average difference in the average easy round performances between the displayed distributions: hiring decisions between workers in the *Information Stage*

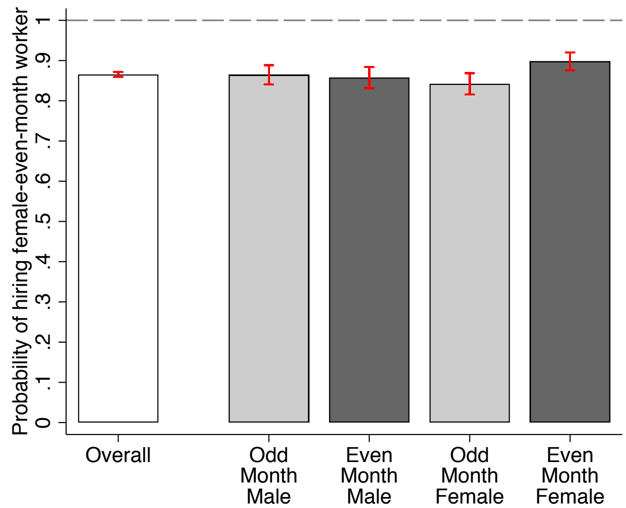
	(1)	(2)	(3)
In-Group	-0.032	-0.037	-0.036
	(0.043)	(0.061)	(0.061)
Gender Treatment		0.001	0.004
		(0.060)	(0.061)
*Gender Treatment*In-Group*		0.009	0.010
		(0.086)	(0.086)
Posterior(easy gap)			0.001
			(0.008)
Posterior(hard-easy gap)			-0.003
			(0.012)
Constant	0.404***	0.404***	0.400***
	(0.030)	(0.041)	(0.044)
Decision FEs	no	no	no
Observations	400	400	400

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making decisions in the *Information Stage* involving displayed distributions that indicate that there is no average difference in easy round performances between the two distributions. ***In-Group*** is an indicator for employers who know they share some demographic characteristic with the female-even-month workers (i.e., even-month employers in the Birth Month treatment and female employers in the Gender treatment). ***Gender Treatment*** is an indicator for employers in the Gender treatment. ***Posterior(easy gap)*** is an employer's posterior belief about the average performance gap in the easy quiz. ***Posterior(hard − easy gap)*** is an employer's posterior belief about the average performance gap in the hard quiz minus the employer's posterior belief about the average performance gap in the easy quiz. All hiring decisions are from the Birth Month and Gender treatments.

Figure A.5: Hiring decisions between workers with same payoff rule but different performances (i.e., the female-even-month worker has a better easy round performance than the male-odd-month worker)



(a) Gender Treatment



(b) Birth Month Treatment

Table A.9: Hiring decisions between workers with different performances (i.e., the female-even-month worker has a better easy round performance than the male-odd-month worker)

	Same payoff rule			Risk in payoff rule only for female-even-month workers		
	(1)	(2)	(3)	(4)	(5)	(6)
In-Group	0.015	0.026	0.023	-0.021	-0.029	-0.033
	(0.016)	(0.025)	(0.025)	(0.020)	(0.027)	(0.027)
Gender Treatment		0.031	0.029		0.024	0.026
		(0.023)	(0.023)		(0.025)	(0.025)
*Gender Treatment*In-Group*		-0.021	-0.022		0.016	0.010
		(0.032)	(0.032)		(0.039)	(0.039)
Posterior(easy gap)			-0.003			-0.011***
			(0.003)			(0.004)
Posterior(hard-easy gap)			0.005			-0.007
			(0.004)			(0.005)
Decision FEs	yes	yes	yes	yes	yes	yes
Observations	4800	4800	4800	24000	24000	24000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the employer-level and shown in parentheses. The results are from ordinary least squares regressions of the probability with which a female-even-month worker is hired over a male-odd-month worker, among employers making hiring decisions between pairs of workers in which the even-month-female worker has a better performance than the male-odd-month worker. ***In-Group*** is an indicator for employers who know they share some demographic characteristic with the female-even-month workers (i.e., even-month employers in the Birth Month treatment and female employers in the Gender treatment). ***Gender Treatment*** is an indicator for employers in the Gender treatment. ***Posterior(easy gap)*** is an employer's posterior belief about the average performance gap in the easy quiz. ***Posterior(hard − easy gap)*** is an employer's posterior belief about the average performance gap in the hard quiz minus the employer's posterior belief about the average performance gap in the easy quiz. Decisions FEs include 12 fixed effects (with one excluded) that capture whether hiring decisions involve a male-odd-month worker versus a female-even-month worker where the female-even-month worker has a better performance on the sports quiz (4 v 5, 4 v 6, 4 v 7, 4 v 8, 6 v 7, 6 v 8) or on the math quiz (4 v 5, 4 v 6, 4 v 7, 4 v 8, 6 v 7, 6 v 8) . Columns 1 - 3 involve hiring decisions where the payoff rule is always the same, while Columns 4 - 6 involve hiring decisions where the payoff rule is risky when hiring female-even-month workers and riskless when hiring male-odd-month workers. All hiring decisions are from the Birth Month and Gender treatments.