Hidden Gender Discrimination

Elisa Macchi* and Claude Raisaro‡

Version: December 23, 2024. Please do not cite without permission

Abstract

We study gender preferences in hiring within three male-dominated sectors in Uganda. We document that women are perceived to excel in trustworthiness — a valued but hard-to-find trait — and an unmet demand for female workers. We design an intervention to match vocationally trained women with firms, building on these key facts. We randomize gender profiles to measure gender differences in hiring preferences and we randomize the provision of monitoring support. We find that both the demand for hiring women and the gender hiring gap vary based on workplace diversity preferences. At the top quintile, managers prefer women over equally qualified men. However, reducing monitoring costs widens the gender hiring gap, particularly among managers with high diversity preferences, indicating hidden gender discrimination rooted in trust perceptions. Monitoring the firm does not impact the gender gap in hiring. Our study implies that gender gaps might be more extensive than previously estimated and that reducing asymmetric information may have unintended consequences for the persistent gender segregation in the labor market in poor countries.

*Brown University (email: elisa_macchi@brown.edu)

†Geneva Graduate Institute (email: claude.raisaro@graduateinstitute.ch)

‡We thank Siwan Anderson, Abhijit Banerjee, Aditi Bhowmick, Lorenzo Casaburi, Rocco Macchiavello, Rohini Pande, Frank Schilbach, David Yanagizawa-Drott and various participants at NEUDC and MIT Behavioral Lunch for helpful discussion and comments, and Josh Bwiira for his outstanding field management and research assistance. The experiments were approved by Mildmay Uganda (0408-2023), and the Uganda National Council for Science and Technology (SS2574ES). The experiment was pre-registered at the AER Registry (AEARCTR-0013698).

1 Introduction

Efforts to enhance women's labor market outcomes have predominantly targeted the supply side. However, these policies may be undermined by demand-side factors, as managers can serve as gatekeepers in male-dominated industries. We demonstrate that when managers value prosocial behavior, a subtle form of discrimination against women emerges, becoming visible only when the value of these traits is diminished or removed. This hidden discrimination dampens the true extent of bias in standard discrimination measures.

How do male managers react to (trained) women (trying to) enter traditionally male sectors? Traditional gender norms suggest that there may be bias, in line with recent evidence (Buchmann et al., 2023; Heath et al., 2024). Yet, women are often perceived as more prosocial (cooperative, reliable, and trustworthy) (Exley et al., 2024), and beliefs about gender differences in behavior, which in turn are likely amplified by traditional norms, play a role in discrimination (Coffman et al., 2021). We investigate this trade-off in Uganda, focusing on three male-dominated sectors: mechanics, welding, and carpentry.

We document key aspects of managerial hiring preferences by gender. First, although only 16.4% of managers employ women, 81% express interest in hiring them, suggesting an unmet demand for female employees.[1] Second, managers, who rate good behavior as the most valued trait in employees, perceive female workers as more trustworthy and reliable than their male counterparts. Third, when asked why they do not hire women, most managers attribute it to a lack of opportunities. Building on these findings, we ask three questions. Are male managers actually interested in hiring women? Is there a gender gap in hiring? Is there evidence of hidden gender discrimination, whereby the gender gap in hiring increases when monitoring costs are reduced?

We conduct a field experiment in Kampala connecting male and female trainees from vocational training institutes (VTI) with 618 firms. We employ Kessler et al. (2019)'s Incentivized Resume Rating paradigm to investigate gender preferences in hiring in an incentive-compatible framework. We build hypothetical resumes from students' information gathered from VTIs enrollment administrative data. To evaluate the interplay of bias and skills, each resume cross-randomizes gender and ability. We select hiring managers looking for trainees and ask them to assess up to 36 hypothetical CVs. Our main outcome

[1] All managers are male. 66% are open to hiring multiple women. Only 7% desire a balanced gender mix. The median firm size is 7 workers.

is the managerial decision to meet the candidate to offer an internship position.[2]

To test for hidden discrimination, we randomize monitoring support and explore the effects on the gender gap in the experiment. In the Monitoring-Trainee arm, we support firms with unannounced audit visits that "discourage new workers from practices such as stealing, dishonesty, disrespect" to reduce monitoring costs. According to the hidden discrimination hypothesis, lower monitoring costs should widen the gender gap by reducing the importance of trustworthiness — a trait managers typically associate with women.

Audit visits may also impact hiring through additional channels that vary by gender. First, via managerial behavior: audits may deter managers inclined toward harassment, with the unintended effect of reducing female hiring. Second, via concerns about safety or harassment: managers who are hesitant to hire women due to fears of harassment by others (clients or co-workers) may feel more reassured by the increased oversight, making them more likely to hire women. To isolate these effects, we introduce a Monitoring-Workplace arm where one third of the sample is randomly assigned to receive unannounced audit visits to "ensure that new workers are safe and treated with respect." The Pure-Control arm is not informed about any visits (business as usual with no visits).

Are managers interested in hiring women? We measure the gender hiring preferences of managers under business-as-usual conditions (Pure Control). We find that 90.1% of managers select at least one female CV in the experiment, and 83.8% select two or more. Of the selected CVs, 42.7% are women.[3] Managers with preferences for a more balanced gender mix select more women: a 10 percentage points increase in the optimal share of women in the firm predicts a 5.7 percentage point increase in the share of women among the selected CVs. Thus, our incentive-compatible experimental evidence confirms a demand for women workers in these male-dominated sectors.

Is there a gender gap in hiring? Resumes are randomly assigned to gender, therefore we can compare managerial preferences for hiring all else equal.[4] In the raw data, on average, a manager asks to meet with 39.7% of the women candidates, while they ask

[2] We also collect beliefs about candidates' work quality and behavior. Managers know that their ratings will determine trainee referrals with desired characteristics. In total, managers make 18'394 evaluations.

[3] On average, managers request meetings with 44% the trainees whose CVs they evaluate.

[4] On top of gender, the CVs include information on: age, marital status, language spoken, nationality, educational achievement, vocational training, GPA in the vocational training class, references, and motivation for entering the male-dominated field.

to meet with 48.4% of the men. These results suggest a gender gap in hiring, yet this gap is unexpectedly small and statistically imprecise, particularly given the high degree of gender segregation in the sectors.

The gender gap in hiring is highly heterogeneous by gender diversity preferences. A stronger preference for a diverse gender mix predicts higher hiring interest in women relative to men, given equal qualifications. The top quintile of this preference distribution, managers whose optimal gender mix is at least 4 women out of 10 workers, positively discriminate in favor of female candidates. The average gender gap in hiring against women in the first four quintiles is 11 percentage points and is statistically significant at the 10% confidence level (p-value 0.042). The gender gap is most pronounced among managers whose ideal worker composition is fully male. By contrast, the gender gap is not differential between high-skill and low-skill candidates. A top GPA significantly increases the hiring for male trainees by 11.9 percentage points (28.7%), and there is no differential effect for women.

Our results support the hidden gender discrimination hypothesis: we find that when preferences for a gender-balanced workforce are based on gendered perceptions of trustworthiness, managers with such preferences shift their hiring practices more strongly against female candidates as monitoring constraints are eased. We test whether the gender gap in hiring widens when we provide monitoring support. We compare the gender gap in hiring decisions between managers assigned to the Monitoring-Trainee arm, with audit support for trainee oversight, and those in the Pure-Control arm. We find that managers in the Monitoring-Trainee arm are less likely to hire women than in the business-as-usual condition, an effect driven by managers who express preferences for a more diverse worker-gender mix. Managers with at least some preference for gender diversity in the workplace are 5.24 percentage points less likely to hire women relative to men in the Monitoring Trainee group compared to the Pure Control group, effectively doubling the observed gender gap (p-value 0.072). Managers in the top 50% of the diversity preferences distribution are 9.2 percentage points less likely to hire women when offered monitoring support (a 2.6 times increase relative to the gender gap in the Pure Control, p-value 0.070).

We do not find evidence in support of harassing behavior by managers or broader concerns for women's safety (stemming from clients or workers) as drivers of employers' hiring decisions. Evaluating the effect of workplace audits (Monitoring-Workplace arm), we find that these visits have no significant effect on the gender gap in hiring (p-value 0.857), with a coefficient of about half the Monitoring-Trainee estimate. We also find no meaningful correlation with preferences for diversity in the worker-gender mix. In the within-subject version of our experiment, we can reject the hypothesis that the gender gap differs between the Monitoring-Trainee arm and the Monitoring-Workplace arm (p-value 0.003).[5]

Our experimental results show that women have a comparative advantage in perceived trustworthiness, a valuable yet difficult-to-find trait for firms, which offsets the preference for male workers. Ignoring hidden gender discrimination would lead to estimating negligible gender differences in hiring against women for over 80% of our sample, and even a preference for women among one-fifth of the firms. However, the gender gap nearly doubles when monitoring costs are reduced, implying that addressing asymmetric information in the labor market may inadvertently disadvantage women.

The experiment is not designed to identify the specific underlying beliefs that drive managers' perceptions of women's trustworthiness—such as selection, lack of outside options, expectations of punishment or socialization. However, we leverage survey data from managers and trainees to examine some of these potential drivers and assess the accuracy of managers' beliefs. We find no evidence that managers' perceptions of women's trustworthiness are influenced by expectations about women having lower outside options or facing harsher punishment for misbehavior. In addition, managers do not systematically view women as more prosocial in general. Instead, using a game designed to test effort and lying behavior under supervision, we find that managers believe incentives and supervision are more likely to influence men's behavior than women's. None of these beliefs are supported by the trainee data, suggesting biased expectations.

Our paper contributes to three main literatures. First, most literature measures discrimination via outcome differences based on observable characteristics (Goldin and Rouse, 2000; Bertrand and Mullainathan, 2004; Bohren et al., 2023; Macchi, 2023; Bertrand and Duflo, 2017). We design an experiment to demonstrate that when the candidate pool is heterogeneous along a relevant but unobservable trait, these standard measures "hide" discrimination. That is, when the decision maker expects members of a discriminated group to outperform the majority group on a key unobservable trait, equal – or even superior – performance of candidates from that group does not rule out discrimination.

[5]For approximately two-thirds of the sample, carpenters and welders, we implemented a within-subject version of our experiment where managers made hiring decisions under varying treatment conditions, with audit visits randomized ex-post. This design was not pre-registered and is therefore not our primary specification, but offers greater statistical power.

Our findings connect with Coffman et al. (2021), showing that beliefs about gender differences in behavior can amplify discrimination, the framework of experimentation over discrimination of Macchiavello et al. (2020), the results of Bohren et al. (2019), when beliefs are biased, a woman needs to produce higher quality output to receive a similar evaluation as a man, and the recent work on systemic discrimination (Kline et al., 2022; Bohren et al., 2022).

We also contribute new insights into gender discrimination by focusing on male-dominated sectors in poor countries.[6] While most empirical research on gender bias in hiring, promotions, and wages is conducted in high-income countries (Bertrand and Duflo, 2017; Delfino, 2024), few studies in low-income settings examine sectors where women are already present. For example, Brown (2023) analyzes the gender bias among school managers toward female teachers. An exception is Buchmann et al. (2023), which highlights a novel form of discrimination, paternalistic discrimination, among managers in Bangladesh in the context of gender neutral but potentially unsafe jobs.

Second, we contribute to the large literature on barriers to female labor force participation and the determinants of gender gaps in the labor market in poor countries. While a significant body of research studies gender inequality in labor markets in poor countries (e.g., Jayachandran, 2015, 2021; Dhar et al., 2019), the role of managers as gate keepers has only been recently highlighted by Buchmann et al. (2023) in relationship to paternalistic discrimination. Moreover, within this literature, social norms (e.g., Bursztyn et al., 2020; Bernhardt et al., 2018), gendered beliefs (e.g., Exley et al., 2024; Feld et al., 2022; Coffman et al., 2021), and stereotypes (e.g., Bordalo et al., 2019; Carlana, 2019), are typically shown to negatively affect women outcomes. We show that managerial beliefs about women's performance can provide a comparative advantage and improve women's labor market outcomes when the "female-typical" trait is relevant and scarce.

Third, our results contribute to the literature on the value of trust for firms in markets characterized by high informational asymmetries (Beaman and Magruder, 2012; Heath, 2018; Bolte et al., 2020; Chandrasekhar et al., 2020; Boudreau et al., 2023). Previous studies highlight that informal hiring technologies, such as network or referral hiring, enhance firm productivity by facilitating monitoring and aligning incentives within the firm. As shown by Alfonsi and De Souza Ferreira (2024), these referral networks are often

[6] This is an important margin for addressing the gender gap globally. Goldstein et al. (2019) documents that women entering male-dominated sectors in poor countries earn, on average, more than those remaining in traditionally female-concentrated sectors.

gender-biased. We highlight an additional layer through which these networks may affect demand for female workers, via reducing the need for monitoring.

2 Setting

In Uganda, female labor participation is around 50% (World Bank), and gender segregation is pronounced, with most occupations almost entirely single-gendered. We focus on three sectors: mechanics, welding, and carpentry. These sectors are emerging urban working-class occupations in both formal and informal settings and are heavily male dominated. We cooperate with vocational training centers (VTI) to gather information on trainees.

2.1 Sample Selection

The study population consists of motor mechanics garage managers, carpenters, and welders located within the metropolitan area of Kampala, Uganda. Managers were selected for participation based on the following eligibility criteria: (i) hiring managers from small and medium-sized firms (below 250 employees), (ii) aged 20 years old or above, (iii) interested in hiring a trainee within a 9-month time window, and (iv) English or Luganda speakers. Field surveyors approached potential candidates at their workplace, following a random walk in the targeted areas, reaching a sample size of 618 firms. Respondents are identified as hiring managers if they have hiring discretion (under our definition, an owner directly involved in hiring is considered as a manager).

2.2 Descriptive Statistics

We interview managers from 317 garages, 190 carpentry firms, and 110 welding firms. Table 1 describes the sample. The firms in our sample are small and medium enterprises, employing an average of 10.7 workers (median: 7 workers). On a typical day, they serve around 9.4 clients, with about 51% being repeat customers. These firms are well-established, having been in operation for an average of 11.1 years (median: 9 years).

Firms are profitable, with an average monthly profit of USD 179 (median: USD 74).[7]

In line with nationally representative data, the firms are strongly male dominated. Most hiring managers (96.4%) are men, with an average of more than 16 years of industry experience. In addition, one in three managers has received some form of vocational training. Approximately 85% of managers are firm owners, although among women managers, only 27% are owners or co-owners. Most managers do not face challenges in retaining good workers. Qualitative data shows that employers retain workers with high wages and good working conditions. Managers bear the responsibility of monitoring workers and state constraints in doing so, with over 85% expressing a desire to monitor more than they currently can. On average, managers spend 10.38 hours at the firm daily, dedicating 2.06 hours to monitoring workers, making it the second highest-ranked activity in terms of time allocation — after technical work and before training workers. 85.1% of managers state they would want to spend more time monitoring workers than they can. 39.7% of managers state that such monitoring constraints are an issue for the productivity of the business.

Most workers are compensated on a piece-rate basis. Workers on probation are typically unpaid (49. 4%) or paid by piece rate (22%). The workforce is also heavily male-dominated. Out of 4,881 currently employed workers, 163 are women. The majority of firms (83.6%) do not employ any women, and the average share of female workers is 0.12 among those that do. On average, 1.65 out of 10 workers have received vocational training, rising to 2.45 if the manager is vocationally trained, and 2.42 if the firm employs any female workers. Many workers are connected to the managers' networks: 41.1% are family members or close friends, and someone in the manager's network has referred 84%.

3 Stylized Facts

In investigating managerial preferences for workers and hiring women specifically, we highlight three key facts.

Behavior and trust as most important traits in workers A majority of managers identify good behavior, honesty, and trustworthiness as the most important traits

[7]For reference, the Ugandan GDP per capita in 2024 is USD 1,023.

for workers, mentioning these qualities nearly twice as frequently as effort and discipline, ability and motivation to learn, and skills (the second, third and fourth traits most mentioned, respectively; Figure 1, panel (a)). Not only are these behavioral traits valued among the most valued, but they are also considered the hardest to find and monitor, as shown in panel B of Figure 1.

Women perceived as better behaved but less skilled Gender-based stereotypes shape managers' perceptions of worker traits. Managers view men as stronger in the ability domain (work quality, learning ability, and effort), while women are seen as superior in trust-related traits, such as good behavior and no stealing, as summarized in panel (c) of Figure 1. There is no gendered perception of cooperation skills.

Managers do worry about the negative productivity consequences of hiring women for their business. In fact, when asked about the consequences of hiring women, 45% of managers state that productivity will not change, while another 45% anticipate an increase in productivity. When asked why, most managers state that women attract customers.

Unmet demand for female workers A majority of managers favor some gender diversity in the workforce at their firm.[8] Specifically, 82.6% of managers prefer at least one woman in a team of ten workers and 67.2% state that the best composition includes at least two women. However, only 5.67% of the managers express a preference for a balanced gender composition, and only 1.13% prefers the majority of women workers.

Despite these preferences, 83.6% of firms currently employ no women, and 94.2% hire at most one woman, suggesting an unmet demand for female workers. Social desirability bias may partly explain the gap between stated demand for female workers and actual hiring, though we find little evidence for this. Most managers do not perceive the goal of our study as encouraging them to hire women, and few are aware of gender-affirmative action policies (e.g., the UNDP "Gender equality seal" for private enterprises in Uganda).

Managers cite the lack of opportunities to hire women as the main reason they do not hire women (mentioned by more than 50% of the respondents). When asked about the share of female workers in their standard applicants pool, managers report that less than

[8] Preferences are elicited through the question, "What is the best gender composition of workers in your firm?" on a scale from "0 out of 10" to "10 out of 10".

5% of applicants are women (0.5 out of 10). We investigate whether this is possibly due to hiring practices primarily based on networks. Indeed, over 80% of employers state they search via their network (friends, family, clients, employees). About 25% of the managers search for workers also in VTIs.

4 Experimental Design

Motivated by these key facts, we design a field experiment to address three questions: Are male managers at all interested in hiring women? Is there a gender gap in hiring, conditional on observables? Is there evidence of hidden gender discrimination, whereby the gender gap in hiring increases when monitoring costs are reduced? Our experimental design randomizes the information about the trainee profiles that managers are asked to evaluate for an internship and the monitoring conditions under which the internship takes place. In what follows, we describe the profile evaluation process and the randomization of both job candidates' characteristics and monitoring conditions.

4.1 Incentivized Resume Rating

Hypothetical profiles. On the supply side, we collect administrative data on 285 enrolled trainees from six partner VTIs, located in the metropolitan area of Kampala. We combine trainee information from the VTI administrative data to create 36 hypothetical resumes by cross-randomizing information about age, gender, marital status, driver's license, language spoken, motivation, education, training, ability, references, and nationality (all Ugandans).[9] Figure 3 illustrates an example of resumes as presented to hiring managers.

Task and incentives. We ask managers to evaluate the profiles of up to 36 candidates during their working time.[10] Managers are informed that the profiles are hypothetical. Following the Incentive Resume Rating paradigm developed by Kessler et al. (2019), we

[9] The information provided in the resumes is chosen based on initial focus groups with a separate pilot sample of 25 managers.

[10] The evaluation process includes pauses after the 8th, 16th, and 24th profiles, dividing the process into four blocks.

inform respondents that, at the end of the study, referrals will be offered from a pool of real candidates associated with our vocational training center partners.[11] Each respondent is guided through the incentive structure before the beginning of the profiles' evaluation.

Outcomes. We elicit two measures of hiring preferences as our main outcomes, plus a set of secondary outcomes to investigate the mechanism. The primary outcome is:[12]

- *Meet*: Do you want us to refer to you a similar worker to start a probation period at your firm? (Yes/No)

The secondary outcomes are:

- *Work Quality*: How would you rate the worker's skills and work quality? Please rate on a scale from 0 to 10, where 0 is very low quality and 10 is very high quality.
- *Behavior*: How would you rate the workers' behavior (trustworthiness and honesty)? Please rate on a scale from 0 to 10, where 0 is not at all trustworthy/honest and 10 is very trustworthy/honesty.
- *Earnings*: What is your best guess of the monthly earnings of this worker a year from now?

4.2 Candidates Gender and Ability Variation

Evaluating the interest in hiring women from the existing pool of applicants considered by managers is not suitable, since the share of female applicants is very limited.[13] Comparing the evaluation of existing female and male candidates does not provide a suitable test for the gender gap in hiring conditional on observables, since the supply of candidates might differ by gender.

[11] The respondents are also informed that the trainees are referred to the firm at the end of their training program.

[12] We also elicited a second primary outcome, the likelihood of extending an offer after the probation period, however, due to a programming error the outcome was only elicited to less than half of the sample.

[13] Less than 5% of the typical candidates are women.

We match each manager with a list of up to 36 job candidate curricula where female and male profiles are randomly assigned in equal proportions.[14] This design allows us to test the interest in hiring women and the difference in hiring by gender. As each manager evaluates either the female or the male version of a given profile, we exploit random assignment to estimate the effect of gender on hiring outcomes all else equal, across managers. To test whether candidates' ability drives the gender gap, we cross-randomize information on gender with information on the candidate's ability. In particular, we randomly assign profiles, with equal probabilities, to two relative performance levels: average GPA (3 out of 5) and high GPA (5 out 5) within their vocational training program. Mirroring the gender dimension, our cross-randomization allows us to evaluate differences across gender and ability in 2×2 design with variation across managers for each profile.[15]

4.3 Monitoring Variation

The hidden gender discrimination hypothesis suggests that the gender gap in hiring widens when monitoring costs associated with mistrust in workers are reduced. Testing this prediction requires comparing the gender gap under business-as-usual conditions with the gap observed when the need to monitor worker behavior is reduced. To achieve this, we introduce random variation in the provision of monitoring support during the internship period for the referred workers who are ultimately hired.

Across-subject Randomization. Managers are randomly assigned into three groups with equal probabilities. All participants are informed before the evaluation process that the research team aims to ensure that "managers and workers will have a positive experience if they end up being matched". In one group, the Pure-Control arm, managers, however, receive no monitoring support, and choices are made under standard business-as-usual conditions.

In the second group — the Monitoring-Trainee arm — receives audit support aimed at overseeing new trainees. Managers are informed that if they hire referred workers, their firm may receive unannounced weekly visits carried out by a member of the training center

[14] We document at baseline that 1) candidates are typically coming from managers' personal network and 2) networks of potential job candidates are male-dominated.

[15] In other words, the cross-randomization allows us to test for beliefs-based vs. preference-based discrimination.

and the research team, with the likelihood of these visits determined randomly through a lottery.[16] The total amount of hours of monitoring expected is calibrated based on the elicited average unmet monitoring needs of managers of two to three hours. Importantly, managers are informed that these visits are designed to monitor and discourage undesirable behaviors by new workers, such as stealing, dishonesty, and disrespect.

Managers may perceive these visits as a signal of lower applicant quality compared to the Pure Control group. To mitigate this concern, managers are explicitly assured that the visits are entirely random and not based on any characteristics of the firm or the workers.[17]

In the third group, the Monitoring-Workplace arm, managers are explicitly informed that the visits are designed to ensure that workers are safe from harassment and treated with respect, without any focus on monitoring behaviors such as stealing, dishonesty, or disrespect. This second type of monitoring is introduced for two key reasons. First, it isolates the hidden discrimination channel — the impact of reducing monitoring costs for new workers — from other potential effects of audit visits on hiring decisions that may vary by gender due to workplace dynamics. Specifically, audits may influence managerial behavior by deterring those inclined toward harassment, which could unintentionally reduce female hiring. Audits may also reassure managers who hesitate to hire women due to concerns about harassment by others, such as clients or coworkers, thereby increasing their willingness to hire women. In the latter case, comparing the evaluations in the Monitoring-Trainee group with the ones in the Pure-Control group would result in an underestimate of the hidden discrimination channel. In the former case, comparing evaluation in the Monitoring-Trainee group relative to Pure-Control group might overestimate the impact of the hidden discrimination channel.

Second, we are interested in evaluating the effect of providing audit support to ensure that new workers face a safe and respectful workplace environment, particularly about potential harassment.[18]

[16] Managers are also informed that the incoming hires are made aware of these visits.

[17] Managers in MT arm are told the following: "We are committed to ensuring that managers and workers will have a positive experience if they end up being matched. Our team members will conduct unannounced weekly visits to some of the firms where workers are placed. Workers will also be informed about the visits. Your firm may be randomly selected to receive such support visits via a lottery. Receiving visits does not reflect any characteristics of the firm or the workers. These visits discourage new workers from practices such as stealing, dishonesty, and disrespect by way of monitoring."

[18] Managers in MW arm are told the following:"We are committed to ensuring that managers and workers will have a positive experience if they end up being matched. Our team members will conduct

Within-subject Randomization. After eliciting evaluation ratings for the first set of 24 resumes under one of the three assigned monitoring conditions, participants are randomly assigned to evaluate 12 additional resumes under one of the remaining two monitoring conditions.[19] We call the first round used in the across-subject design "round one" and the subsequent set of evaluations "round two".[20] This design serves three purposes. First, it allows us to estimate treatment effects using within-subject variation, providing both greater statistical precision and a benchmark for the across-subject design. Second, it enables a skewed mass assignment of monitoring supports toward those conditions that have the largest positive impact on reducing the gender gap and increasing labor demand. Third, it allows the calculation of individual-level differences in evaluations between two monitoring conditions, creating a personalized measure of the treatment effects.

4.4 Sample Selection

In the main analysis, we restrict the sample to participants who pass the attention checks on the information provided on the monitoring support and exclude cases where at least one evaluation block shows no variation in the primary outcomes. The main analysis sample consists of 524 managers, each representing a different firm. Importantly, the sample selection procedure was preregistered and does not predict treatment assignment.

4.5 Randomization Balance

Table 2 displays the balance in the observable sociodemographic and work characteristics of the hiring managers and the characteristics of the firms in the sample under different

unannounced weekly visits to some of the firms where workers are placed. Workers will also be informed about the visits. Your firm may be randomly selected to receive such support visits via a lottery. Receiving visits does not reflect any characteristics of the firm or the workers. The visits are to ensure that the new workers are safe (not harassed) and treated with respect. These visits cannot discourage new workers from practices such as stealing, dishonesty, disrespect by way of monitoring."

[19] We implemented the within-subject design for all managers except those in the motor-mechanics sector. This extension was introduced after the first phase of data collection, which focused exclusively on one sector, to improve the statistical power of the regression analysis.

[20] For example, if a manager is initially assigned to the pure control in the across-subject design, they are subsequently assigned to either the MT or MW condition with equal probability. In summary, study participants are assigned to one of the following pairs sequences with equal probabilities: PC-MT, PC-MW, or MT-MW.

monitoring conditions.[21] While most of the baseline variables are balanced, a few variables are unbalanced across treatment arms. Four out of 43 coefficients are statistically significantly different at the 10% level when comparing Pure Control with Monitoring Trainee, the results are robust to specifications with and without controls for unbalanced variables.

5 Empirical Strategy

This section outlines our empirical strategy and is structured around our main research questions.

Are Managers Interested in Hiring Women? We evaluate managers' interests in hiring women when presented with out-of-network vocationally trained candidates in an incentive-compatible setting. The IRR setup ensures incentive-compatible responses with respect to hiring interest, as study participants are seeking new workers and have a vested interest in selecting suitable candidates.

Our descriptive analysis focuses on managers assigned to the Pure Control group in the across-subject design, given that this represents the business-as-usual condition. We measure the average and underlying distribution of the share of women resumes selected out of the total number of resumes selected by managers. As a sanity check, we investigate the correlation between stated preferences for diversity in the workplace and hiring choices.

Is There a Gender Gap in Hiring? To test for gender differences in hiring we examine how hiring outcomes are impacted by the gender of jobseekers. We restrict our baseline sample to firms to which no monitoring support is provided. We first estimate the following regression model:

$$\text{Meet}_{ijs} = \beta_0 + \beta_1 \text{Female}_{ij} + \delta_i + \sigma_s + u_{ijs} \tag{1}$$

[21] Overall attrition rates are low as the intervention was implemented within a one-time survey. The three respondents who did not complete the survey were replaced by another firm manager in the same neighborhood. We do not observe attrition in rating the candidates' profiles, as all managers who completed the survey evaluated the entire set of proposed resumes.

where Meet^k_{ij} denotes manager j's interest in meeting a profile with characteristics of CV i in strata s.[22] Female_{ij} is a dummy for CV i being associated with a female candidate to manager j. The coefficient β_1 captures the gender bias in the form of the differential effect of female as opposed to male candidates on outcome Y. δ_i σ_s are, respectively, profile and strata fixed effects. We cluster standard errors at the manager level.

Given the observed heterogeneity in workplace gender diversity preferences in our sample, we investigate heterogeneity in the gender gap by estimating the following regression model for the Pure Control group:

$$\begin{aligned}\text{Meet}_{ijs} = \beta_0 + \beta_1\text{Female}_{ij} + \beta_2\text{GenderDiversity}_j + \\ \beta_3\text{Female}_{ij} \cdot \text{GenderDiversity}_j + \delta_i + \sigma_s + u_{ijs}\end{aligned} \tag{2}$$

where GenderDiversity_j represents manager j's preference for gender diversity, defined as the preferred proportion of women in their ideal workforce composition. The main coefficient of interest is β_3, which captures the differential interest in hiring women relative to men for managers with 1 standard deviation increase in the preferred share of female workers.

To further investigate the drivers of gender differences in hiring, we focus on examining the impact of ability on hiring decisions and its interaction with gender. This test allows us to investigate to what extent managers engage in differential hiring behavior by gender because of statistical versus taste-based discrimination. If managers are statistically discriminating against women, we should expect that gender differences in hiring are lower for CVs randomly assigned to a high GPA. We estimate the following regression on the subset of resumes in the Pure Control group:

$$\begin{aligned}\text{Meet}^k_{ijs} = \beta_0 + \beta_1\text{Female}_{ij} + \beta_2\text{HighGPA}_{ij} + \\ + \beta_3\text{Female}_{ij} \cdot \text{HighGPA}_{ij} + \delta_i + \sigma_s + u_{ijs}\end{aligned} \tag{3}$$

where HighGPA_{ij} is a dummy indicating whether CV i presented to manager j has a top score (5 out 5) within their vocational training program as opposed to an average grade (3 out of 5). The coefficient of interest is β_3, which captures the differential effect of gender

[22] Randomization is stratified at the sector, survey date, enumerator, and resume-evaluation order (three blocks of eight CVs each) level.

by ability on the hiring interest.

Is There Hidden Gender Discrimination? We test for hidden gender discrimination by examining whether reducing monitoring costs associated with new potential hires widens the gender gap in hiring, disadvantaging female candidates.

In addition, the hidden discrimination hypothesis suggests that if women's hiring is driven by their perceived trust advantage — managers viewing women as more trustworthy — then hidden discrimination (the impact of monitoring on the gender gap) should be stronger among managers with stronger preferences for women at baseline. We first estimate the following regression model on the full sample of managers and on managers whose preferences do not favor an all-male workforce:[23]

$$\begin{aligned} Y^k_{ijs} = \beta_0 &+ \beta_1 \text{Female}_{ij} + \beta_2 \text{MT}_j + \beta_3 \text{MW}_j \\ &+ \beta_4 \text{Female}_{ij} \cdot \text{MT}_j + \beta_5 \text{Female}_{ij} \cdot \text{MW}_j + \delta_i + \sigma_s + u_{ijs} \end{aligned} \tag{4}$$

where Y^k_{ij} is the outcome k of manager's j evaluation of candidate i. MT_j and MW_j are dummies for being assigned to the Monitoring-Trainee arm and Monitoring-Workplace arm respectively, while the Pure-Control arm is the omitted group. We use the above regression model for all pre-registered outcomes.

Our coefficient of interest is β_4, which captures differences between female and male candidates in the impact of Monitoring-Trainee (MT) support relative to Pure-Control. We interpret a negative coefficient (i.e. an increase in the gender gap against women) as evidence supporting the hidden discrimination hypothesis. We expect a negative estimate of β_4 for the sub-sample of managers who do not prefer an all-male workforce. We are also interested in the β_5 coefficient in itself, as it captures the gender differences (against women) in the impact of offering monitoring support to ensure a safe working environment for the new workers.

The experimental design also allows us to exploit within-subject variation in monitoring

[23] In contrast, we expect the reduction in monitoring costs to increase overall labor demand by mitigating asymmetric information during hiring. Ex-ante, it is unclear which effect dominates, making this a key empirical question. Specifically, these two predictions highlight an equity-level trade-off: while reducing monitoring costs may boost overall labor demand, this increase may predominantly benefit male candidates, exacerbating pre-existing gender disparities in hiring.

support, which provides us with a highly powered regression model. Notice that We estimate the causal effect of offering different types of monitoring support – mirroring the across-subject design estimation – by estimating the following regression model:

$$
\begin{aligned}
Y^k_{ijsr} = \beta_0 + \beta_1 \text{Female}_{ijr} + \beta_2 \text{MT}_{jr} + \beta_3 \text{MW}_{jr} \\
+ \beta_4 \text{Female}_{ijr} \cdot \text{MT}_{jr} + \beta_5 \text{Female}_{ijr} \cdot \text{MW}_j + \delta_i + \mu_j + \sigma_s + u_{ijsr}
\end{aligned} \tag{5}
$$

where Y^k_{ijsr} is the outcome k for profile i evaluated by manager j in strata s at round $r \in \{1, 2\}$. Round one corresponds to the evaluation of the first 24 resumes, while round two corresponds to the evaluation of the subsequent set of 12 resumes. MT_{jr} and MW_{jr} are dummies indicating the random assignment of manager j to the Monitoring-Trainee arm and Monitoring-Workplace arm in round r. μ_j are the manager fixed effects. Similarly to equation 4, our main coefficient of interest is β_4, which measures the gender differences in the impact of reducing monitoring costs associated with mistrust in candidates. We are also interested in β_5 coefficient, which estimates the impact of providing support to monitoring the workplace, and allows us to isolate the same alternative channels as in the across-subject design analysis. Mirroring the analysis in the across-subject design, we also estimate equation 5 for managers whose preferences do not favor an all-male workforce.

6 Main Results

6.1 Demand For Female Workers

The experimental findings align closely with the descriptive evidence, showing a demand for female candidates when we provide managers with the opportunity to hire women. Focusing on the control group, the business-as-usual condition, 90.1% of firms selected at least one female candidate, with an average women selection rate of 42.7% and an almost gender-balance median (44.5%) as shown in panel (a) of Figure 4.

The distribution is highly heterogeneous, however, with about 12.4% of the managers never selecting any female trainees. Managers' lack of interest in hiring women does not seem to be explained by productivity considerations. One of the managers who did not select women reported that having more women at the firm would negatively impact the productivity of the business. The lack of interest is also not explained by the current gender composition of the firm, as none of these firms hires women.

Managers' gender diversity preferences drive the demand for female employees, as shown in Figure 4, panel (b). For every 10 percentage-point increase in the manager's ideal proportion of female workers, the share of female candidates selected for meetings rises by 5.8 percentage points.

This correlation between the stated preferences for diversity and actual hiring decisions validates the significance of these beliefs and choices. Interestingly, 23 managers (11% of the PC group) selected women in the incentive-compatible exercise despite having no stated preference for diversity (ideal mix of 10/10 men). Excluding these managers from the sample does not significantly change the overall demand distribution for women: the median share of female trainees out of selected trainees is 50.1%. When asked to explain their inconsistent behavior, less than one-fourth of managers said they would like to "give [women] a chance"; the others mentioned that women are more honest and easier to monitor.

6.2 Gender Differences in Hiring

Next, we investigate whether managerial evaluations result in gender differences in hiring interest, conditional on observables. This analysis focuses on the status quo condition, the control group sample.

We find that women are about 8.7 percentage points (17.9%) less likely to be selected for a meeting — as compared to men with identical resumes. Figure 5 illustrates the gap for the primary outcome. While this indicates a gender gap against women in hiring, the gap is surprisingly small given the extreme segregation in this context. This relatively small difference suggests an underlying demand for female employees. Column 1 of Table 3 reports estimates for the regression model 1. We document the same results: all else equal, women are 8.64 percentage points less likely to be selected relative to a control mean of 50.4% (p-value 0.00).

The gender gap in hiring is heterogeneous by diversity preferences, as shown in panel (a) of Figure 6. Managers in the bottom quintile of diversity preferences — those who prefer a 100% male workforce — are significantly more likely to hire men over equally qualified women. In contrast, managers in the top quintile, who favor a 40%–60% share of women, are significantly more likely to hire women over equally qualified men. Column 2 of Table 3 reports the estimates for the regression model 2: on average, a 1 standard deviation increase in the preferences for diversity in the workplace is associated with a 9.6 percentage point reduction in the gender gap in hiring (p-value 0.000).

The gender gap is not influenced by the skill set of the candidates. As a first sanity check we show that ability matters for selection in the raw data: When the profile is associated with a higher GPA, the likelihood that a male candidate is selected increases significantly and it is not differential by gender: column (3) of Table 3 shows estimates for the regression framework 3: being assigned to a top GPA score increases the likelihood of being selected by 11.9 percentage points (28.7%, p-value 0.000) among men, with no differential effect by gender (p-value 0.62). As robustness, we also check that the gender gap is not driven by other measures of quality or skills, such as having a certification. Overall, this suggests that beliefs about skills, in the accurate statistical discrimination sense, are not masking gender differences in hiring.

Overall, these results highlight that, while some managers hold strong preferences against hiring women — a pattern expected in a predominantly male sector — the majority seem to exhibit relatively low preference for men versus women, all else equal, and a non-negligible portion of managers appears to positively discriminate women over men.

6.3 Gender Gap With and Without Monitoring Support

One interpretation of the minimal differences in hiring by gender is that they reflect low levels of gender bias. However, our descriptive evidence suggests that women may be preferred due to their comparative advantages in trustworthiness and reliability. The value of trust could be especially pronounced because managers are matched with trainees outside their existing networks, where such attributes are highly valued.

To test for this, we investigate to what extent gender differences in hiring depend on the monitoring costs faced by employers. Specifically, we compare managers who are randomly assigned to receive monitoring support through unannounced weekly visits with those operating under business-as-usual conditions. Because most managers state to be monitoring-constrained, we expect these visits to meaningfully affect the perceived monitoring costs of managers. Our "hidden gender discrimination" hypothesis implies that the gender gap should increase when monitoring costs are lower.

The results show that the gender gap in hiring significantly widens when managers face fewer monitoring constraints due to our monitoring support. In the regression model outlined in Equation 4, managers in the Monitoring-Trainee arm are 12.04 percentage points less likely to hire a female trainee compared to a male trainee with identical characteristics (Table 4, Column (1)). Although this effect is economically substantial, representing a 39.1% increase relative to the control group, it is not statistically significant (p-value = 0.260). However, in our more statistically powerful within-subject specification, involving 77% of the sample, the gender gap in hiring increases by 89% (gender gap coefficient: -0.1613, p-value 0.000) when comparing Monitoring-Trainee arm relative to the business-as-usual condition (Table 4, column (4)).

We expect hidden gender discrimination to have more bite with managers who show some preferences for hiring women. Thus, we estimate the effect of monitoring support excluding managers who state 100% male workplace preferences. Estimates reported in column (2) of 4 show that Monitoring-Trainee treatment increases the gender gap against women by 10% relative to Pure-Control group (p-value 0.072); the effects are consistently stronger and more precisely estimated in the within-subject design (Column (4)).

In the heterogeneous treatment effect framework, Figure 6, panel (b), the increase in the gender gap in hiring against women is driven by managers in the top quintile of the preferences for diversity distribution. This result is consistent with trust-based preferences for female workers (i.e. women are perceived as more trustworthy) and with the hypothesis that reducing monitoring costs increases the gender gap if mistrust in workers drives these costs.

The within-subject results suggest that the negative impact of monitoring support on the gender gap against women is 2.2 times larger than the gap estimated when including managers who prefer an all-male workforce.[24]

[24] The fact that about 20% of the sample does not respond to the monitoring incentives, while meaningful within our framework, was not anticipated, as very few participants in the pilot exhibited such extreme preferences for gender segregation in the workforce. This reduces the power of our between-subject experimental test compared to the pilot-based power calculations and makes the within-subject version of the experiment more relevant.

6.4 Unintended Effects of Monitoring Support

Our audit visits are intended to help the manager monitor the workers, but they may have unintended consequences which can confound the mechanism identification, especially if audit visits differentially impact the hiring of men and women through channels other than trust.

There are reasons to believe that this may be the case. For example, if managers intend to harass female workers, monitoring visits could deter such behavior and reduce the inclination to hire women. Instead, if managers are paternalistic — as in the context of Buchmann et al. (2023) — and have safety concerns regarding interactions with other workers or clients, they might become more inclined to hire women when provided with managerial support. The former scenario could account for some or all of the changes in the gender gap that we previously attributed to the decreased perceived trustworthiness of female candidates. The latter scenario, however, might lead us to further underestimate the overall extent of gender discrimination.

We test whether monitoring support not focused on workers also increases the gender hiring gap. We find that in neither of our specifications, the effect of offering monitoring support is statistically significant. In the between-subject design of the experiment, managers in the Monitoring-Workplace treatment exhibit a small and not statistically significant 0.58% increase in the hiring gender gap against women compared to the control group (p-value 0.857). In the within-subjects specification, which offers higher statistical power, the sign of the is reversed and is also not statistically different from zero (+0.035 increase, p value 0.324). In either model, monitoring the firm does not have a stronger impact on the gender gap when restricting to the sample of employers with some preferences for diversity in the workplace (columns (2)-(3) and (5)-(6)).

The high-powered within-subject design allows us to reject the hypothesis that the effects of the Monitoring-Trainee treatment and the Monitoring-Workplace treatment are equal (Table 4, column (4)).[25]

Thus, audit visits alone do not influence the gender gap; their effect occurs only when they reduce monitoring costs for workers. Therefore, these findings reject the hypothesis that there are unintended consequences of monitoring on the gender gap and confirm our

[25]Results are qualitatively the same and stronger in magnitude for the subsample of managers with preferences for mixed-gender workforce, as reported in 4, columns (2)-(3) and (5)-(6).

evidence of “hidden” gender discrimination.

7 Why Women’s Trust Advantage?

While our findings suggest that women may possess a comparative advantage in trustworthiness, several open questions remain regarding the nature and origins of these advantages. First, it is unclear whether women are inherently more trustworthy or if trustworthiness is merely perceived. In the U.S. context, Exley et al. (2024) demonstrates that people perceive women as more prosocial, although this belief is inaccurate. Second, trustworthiness could be attributed to inherent traits influenced by socialization processes or external factors that promote cooperative and reliable behavior (Macchiavello, 2022). These factors are likely to play significant roles in low-income country settings. Societal expectations and gender norms may shape perceptions of trustworthiness (Jayachandran, 2021), while limited job market opportunities for women can drive the development of reliable behaviors. Understanding the accuracy of managers’ beliefs is crucial for assessing the implications for firms. Furthermore, identifying the determinants of any gender differences can inform policies to improve equality. Notably, the source of trustworthiness (if any), and the managerial beliefs on the existence and source of trustworthiness may not be aligned. In this section, we investigate managerial beliefs about gender differences, and actual gender differences in opportunity, behaviors and preferences to shed light on the drivers of women’s comparative advantage.

7.1 Managerial Beliefs and Misperceptions

Gender Differences in Outside Options Managers believe that male and female candidates with identical profiles will earn different salaries one year after graduation. Specifically, managers estimate that women will earn approximately 11.8% less than equally qualified men, suggesting that women are perceived to have fewer outside options in these male-dominated markets. However, the perceived extent of the gender pay gap is negatively correlated with the preferences for diversity as shown in Figure 7, and in turn the demand for female trainees. Therefore, beliefs about outside options are unlikely to account for perceptions of trustworthiness. Interestingly, our trainees survey data reveal that female trainees have similar earnings expectations and beliefs about likelihood of finding a job as their male counterparts. Thus, we conclude that beliefs about outside options are also unlikely to drive any actual gender-based behavioral differences among trainees.

Gender Differences in Punishment One might hypothesize that if women are found to be misbehaving, they could be punished more harshly, perhaps because such behavior is unexpected or for other reasons. Consequently, this could lead women to exhibit better behavior in anticipation of stronger consequences. However, our survey data does not support this hypothesis. When asked to predict the likelihood of punishing a misbehaving female worker versus a male one, managers estimated that men would be punished 84.8% of the time, while women would be punished only 70.6% of the time. This indicates that women are 14.2% less likely to be punished than men. Thus, concerns about harsher punishment for women do not align with managerial perceptions.

Gender Differences in Social Preferences Managers were asked to estimate how male and female trainees agree with various statements about social preferences[26] Managers perceive that female vocationally trained workers are generally more likely to be trusting and to exhibit positive reciprocity, whereas men are viewed as more risk-prone (Figure 8). Traits such as patience and altruism are considered gender-neutral. However, when comparing managerial beliefs with trainee data, we find that female trainees are just more likely to display positive reciprocity and altruism. There are no significant differences in risk-taking behavior or the likelihood of trusting male and female trainees. Overall, we derive two takeaways from these results. First, beliefs about trustworthiness do not merely reflect broader prosociality beliefs. Second, there are significant misperceptions regarding gender traits.

Gender Differences in Behavioral Response to Supervision We designed an experimental game to assess how supervision influences behavior, specifically focusing on effort and instances of lying or misreporting, which we use as inverse indicators of the

[26]We use measures of economic preferences following Falk et al. (2018). Altruism: "How willing are you to give to good causes without expecting anything in return?"; Risk-taking: "How willing or unwilling are you to take risks?"; patience: "How willing are you to give up something beneficial for you today to benefit more from that in the future?"; positive reciprocity "When someone does me a favor, I am willing to return it."; negative reciprocity "How willing are you to punish someone who treats you unfairly, even if there may be costs for you?"; and trust "I assume that people have only the best intentions."

need for monitoring.

Vocationally trained male and female workers are assigned to two randomly assigned conditions: supervision and no supervision. In the no-supervision condition, workers were instructed to perform up to 12 tedious but simple tasks without any oversight and were paid a fixed rate, independent of the number of tasks completed. Consequently, the actual number of tasks each worker completed could not be verified ex post. In contrast, the supervision condition used a piece-rate compensation system, where supervisors verified the number of completed tasks and paid workers accordingly. On the managers' side, we ask managers to predict the average number of tasks completed by male and female workers in each condition. This design allowed us to assess how supervision affects effort and managerial perceptions of worker performance.

In addition to the basic setup, the trainees had the opportunity to misreport their performance by claiming to have completed more tasks to earn 50% higher pay. We designed the task to detect cheating, although the trainees were unaware of this feature. In the no-supervision condition, however, we truthfully could not link reported behavior to individual trainees; instead, we could only measure the average misreporting rate at the group level. Consequently, misreporting was incentive compatible in the no-supervision condition since individual task completions could not be verified. Finally, we asked managers to predict the likelihood of misreporting of male and female vocationally trained (VT) workers under both supervised and unsupervised conditions.

Managers expect VT male workers to complete fewer tasks than VT female workers when unsupervised. However, expected gender differences in task completion disappear when workers are supervised, as illustrated in Figure 9. Because these beliefs are primarily held by managers who have preferences for diversity, expectations of effort under supervision are likely to partly explain the demand for female trainees. Our comparison with the workers shows that these beliefs are not aligned with actual trainees' behavior. Men complete more tasks than women under both conditions, and there is basically no effect of outcome's verifiability. Generally speaking, managers are also very pessimistic on the general trainee effort and optimistic about the value of monitoring/supervision on effort.

Looking at lying behavior, 10, managers expected male VT workers to misreport 24.7% more than female VT workers when unmonitored, reflecting beliefs about women's higher trustworthiness. Under supervision, however, the predicted average misreporting of male workers is reduced by half (p-value 0.000). This indicates that while managers anticipated greater misreporting among men without supervision, they expect monitoring to effectively curb this behavior. These qualitative results further support our experimental evidence and the mechanism of hidden gender discrimination: managers' beliefs about the effects of supervision on work effort and behaviors such as lying or misreporting appear to be the key drivers of the observed changes in the gender gap under different monitoring conditions. Again, these beliefs are not aligned with actual trainee behavior. Women are less likely to misreport than men under both supervised and unsupervised conditions.

7.1.1 Gender Differences in Skills

Data from manager and trainee surveys, summarized in Figure 11, reveal that managers expect gender-based discrepancies in trainee technical abilities across various sectors in line with the general beliefs of the manager about technical ability by gender (Figure 1, panel (c)); however, these perceptions appear incorrect.

We elicit specific beliefs about female trainees' ability to perform a key technical tasks — such as performing an oil change in mechanics or ensuring a smooth finish on wooden surfaces in carpentry — compared to their male counterparts. As shown in Figure 11, managers believe that a lower share of female trainees can perform key tasks compared to male trainees. In contrast, trainee self-assessments indicate no significant difference in the ability to perform these tasks based on gender.

8 Other Explanations of Managerial Hiring Choices

Various additional preferences and contextual factors may influence managerial hiring decisions. In this section, we briefly discuss the main ones and their relationship with our results.

8.1 Customer Discrimination

Managers' concerns about client discrimination could potentially contribute to the gender gap in hiring. However, such concerns are unlikely to result in different hiring preferences for women under monitoring versus unmonitored conditions, except if managers believe that audits disproportionately improve the work quality of female employees. We do not have evidence of that this would lead to increased hiring of women when monitoring is present. Qualitatively, employers do not report fear of customer discrimination, they rather expect women can attract clients.

8.2 Other Workers Do Not Want Women

Our analysis indicates that concerns about discrimination by colleagues or resistance to having women in the workplace are minimal. Specifically, when managers were asked how many workers would accept a woman as a colleague, 56.71% reported that all workers would accept a female colleague. The average expected share of workers who would accept a woman is 79.8%. These findings suggest that negative sentiments towards having women around do not play a significant role in managerial hiring decisions within our sample.

8.3 Gender Norms

We assessed broader gender norms using measures from the World Value Survey. First, approximately 85.9% of managers agree with the statement, "Women should have the freedom to work outside of the home." However, they predict that only 69% of others share this view, indicating pluralistic ignorance—a phenomenon commonly documented in the literature (Bursztyn et al., 2020). Second, 82.8% of managers agree with the statement, "The government and companies should give priority to women when hiring for lead positions," yet they estimate that only 65.54% of others hold the same belief. Finally, 61.5% of managers agree with the statement, "A preschool child is likely to suffer if his or her mother works," and their prediction that 67% of others agree reflects a more accurate assessment of peers' views.

8.4 Experimenter Demands

Experimenter demands might lead managers to hire women to please the experimenters. However, such demands would likely reduce our ability to detect an effect of monitoring on the gender gap by increasing the share of women hired. We do not find evidence of this happening. Indeed, these effects are expected to be stronger in within-subjects designs than in between-subjects designs. Therefore, we should be less likely to detect the monitoring effect in the within-subjects condition. However, our results show the opposite.

To directly check for experimenter demands in the flavor of De Quidt et al. (2018), we asked managers: “In your opinion, would it be better for us if you hired men, women, or does it not matter?”. Most managers state that we would be indifferent. The second most common response, chosen by 23.14% of managers, was that we would prefer they hired men. Only 8.7% thought we would prefer they hired women. These findings suggest that most managers do not feel pressured to hire women to satisfy experimenter expectations.

9 Why Does Gender Segregation Persist?

Given our evidence that women possess a comparative advantage in trustworthiness, it is surprising that the workforce remains extremely segregated by gender. Two factors may contribute to the persistence of segregation.

First, there appears to be a genuine scarcity of women in these industries, as suggested by managers’ perception that only about 4% of applicants are women. However, vocational training centers do train women, suggesting that the limited presence of female candidates in the applicant pool is likely because most managers do not typically hire through VTIs. Our data corroborates this, showing that over 75% of respondents rely exclusively on referrals from other managers, workers, family members, and similar sources. Recent work in the same context indicates that these referral processes can be gender-biased (Alfonsi and De Souza Ferreira, 2024).

Nevertheless, this factor cannot fully explain the persistence of segregation observed, with only approximately 3.4% of workers hired being women, unless one assumes some level of bias. If rational managers truly valued trustworthiness, they would adjust their hiring practices to increase the representation of women accordingly (e.g., by searching in VTIs or directly training female workers).

Persistent segregation implies that managers employ other methods to mitigate concerns related to monitoring workers, which may be particularly preferred in the presence of bias. Indeed, the literature (see Heath, 2018; Chandrasekhar et al., 2020), often attributes the practice of hiring through networks to an effort to reduce monitoring costs. Thus, what our results suggest is that as long as managers can access trusted men via their network, they do not need to engage in the effort of searching for female candidates, perpetuating the existing gender imbalance in the workplace.

10 Conclusions

This study explores hidden gender discrimination in the hiring practices of male managers within Uganda's male-dominated sectors. We find that women possess a comparative advantage in trustworthiness — which leads to an underestimation of the true extent of the hiring gender gap. Importantly, the perception of women's prosociality is not exclusive to developing countries. Additionally, structural factors such as limited outside options may cause other minority groups to be viewed as more prosocial by firms. Consequently, the implications of this study could extend to various contexts involving different minorities, indicating that gender and minority gaps may be more pervasive than previously recognized. This resonates with recent literature on the dynamics of discrimination and systemic bias (e.g., Bohren et al., 2022) and suggests that audit interventions, while methodologically appealing for their simplicity, may fail to capture the nuanced mechanisms of discrimination.

Our research carries significant policy implications. One notable insight is the potential impact of shifting from referral-based hiring systems to more structured and formalized recruitment processes, as observed in larger firms. Such transitions can enhance opportunities for women to enter the labor force by broadening the pool of candidates beyond existing managerial networks. However, this transition may also go along with improved ease of monitoring and control, which could diminish the trustworthiness advantage that women hold. Our study suggests that addressing asymmetric information in the labor market may have the unintended consequence of increasing gender segregation. More research is needed to qualify these trade-offs to design effective interventions aimed at promoting gender equality in the labor market in poor countries.

References

Alfonsi, Livia and Pedro De Souza Ferreira, "Whom would you rather work with? Uncovering the Referral Gender Gap," 2024.

Beaman, Lori and Jeremy Magruder, "Who gets the job referral? Evidence from a social networks experiment," *American Economic Review*, 2012, *102* (7), 3574–3593.

Bernhardt, Arielle, Erica Field, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer-Moore, "Male social status and women's work," in "AEA Papers and Proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 363–367.

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

_ **and Sendhil Mullainathan**, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American economic review*, 2004, *94* (4), 991–1013.

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg, "The dynamics of discrimination: Theory and evidence," *American economic review*, 2019, *109* (10), 3395–3436.

_ **, Kareem Haggag, Alex Imas, and Devin G Pope**, "Inaccurate statistical discrimination: An identification problem," *Review of Economics and Statistics*, 2023, pp. 1–45.

_ **, Peter Hull, and Alex Imas**, "Systemic discrimination: Theory and measurement," Technical Report, National Bureau of Economic Research 2022.

Bolte, Lukas, Nicole Immorlica, and Matthew O Jackson, "The role of referrals in immobility, inequality, and inefficiency in labor markets," *arXiv preprint arXiv:2012.15753*, 2020.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Beliefs about gender," *American Economic Review*, 2019, *109* (3), 739–773.

Boudreau, Laura, Julia Cajal-Grossi, and Rocco Macchiavello, "Global value chains in developing countries: a relational perspective from coffee and garments," *Journal of Economic Perspectives*, 2023, *37* (3), 59–86.

Brown, Christina, "Understanding Discrimination by Managers," 2023. Unpublished manuscript.

Buchmann, Nina, Carl Meyer, and Colin D Sullivan, "Paternalistic discrimination," Technical Report, Working Paper 2023.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived social norms: Women working outside the home in Saudi Arabia," *American economic review*, 2020, *110* (10), 2997–3029.

Carlana, Michela, "Implicit stereotypes: Evidence from teachers' gender bias," *The Quarterly Journal of Economics*, 2019, *134* (3), 1163–1224.

Chandrasekhar, Arun G, Melanie Morten, and Alessandra Peter, "Network-based hiring: Local benefits; global costs," Technical Report, National Bureau of Economic Research 2020.

Coffman, Katherine B, Christine L Exley, and Muriel Niederle, "The role of beliefs in driving gender discrimination," *Management Science*, 2021, *67* (6), 3551–3569.

Delfino, Alexia, "Breaking gender barriers: Experimental evidence on men in pink-collar jobs," *American Economic Review*, 2024, *114* (6), 1816–1853.

Dhar, Diva, Tarun Jain, and Seema Jayachandran, "Intergenerational transmission of gender attitudes: Evidence from India," *The Journal of Development Studies*, 2019, *55* (12), 2572–2592.

Exley, CL, OP Hauser, M Moore, and JH Pezzuto, "Believed gender differences in social preferences," *Quarterly Journal of Economics*, 2024.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global evidence on economic preferences," *The quarterly journal of economics*, 2018, *133* (4), 1645–1692.

Feld, Jan, Edwin Ip, Andreas Leibbrandt, and Joseph Vecci, "Identifying and overcoming gender barriers in tech: A field experiment on inaccurate statistical discrimination," 2022.

Goldin, Claudia and Cecilia Rouse, "Orchestrating impartiality: The impact of "blind" auditions on female musicians," *American economic review*, 2000, *90* (4), 715–741.

Goldstein, Markus, Paula Gonzalez Martinez, and Sreelakshmi Papineni, "Tackling the global profitarchy: Gender and the choice of business sector," *World Bank Policy Research Working Paper*, 2019, (8865).

Heath, Rachel, "Why do firms hire using referrals? Evidence from Bangladeshi garment factories," *Journal of Political Economy*, 2018, *126* (4), 1691–1746.

_ , Arielle Bernhardt, Girija Borker, Anne Fitzpatrick, Anthony Keats, Madeline McKelway, Andreas Menzel, Teresa Molina, and Garima Sharma, "Female labour force participation," *VoxDevLit*, 2024, *11* (1), 1–43.

Jayachandran, Seema, "The roots of gender inequality in developing countries," *economics*, 2015, *7* (1), 63–88.

_ , "Social norms as a barrier to women's employment in developing countries," *IMF Economic Review*, 2021, *69* (3), 576–595.

Kessler, Judd B, Corinne Low, and Colin D Sullivan, "Incentivized resume rating: Eliciting employer preferences without deception," *American Economic Review*, 2019, *109* (11), 3713–3744.

Kline, Patrick, Evan K Rose, and Christopher R Walters, "Systemic discrimination among large US employers," *The Quarterly Journal of Economics*, 2022, *137* (4), 1963–2036.

Macchi, Elisa, "Worth your weight: experimental evidence on the benefits of obesity in low-income countries," *American Economic Review*, 2023, *113* (9), 2287–2322.

Macchiavello, Rocco, "Relational contracts and development," *Annual Review of Economics*, 2022, *14* (1), 337–362.

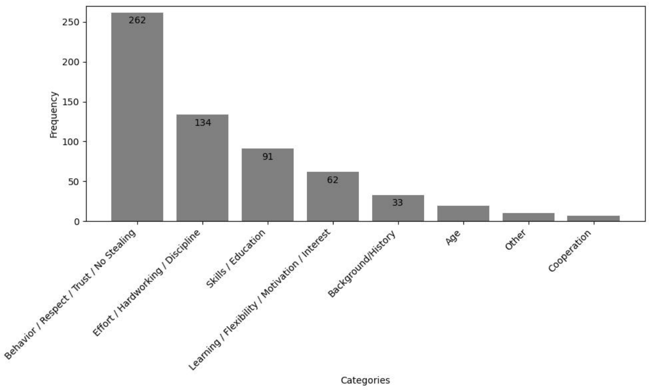
_ , Andreas Menzel, Atonu Rabbani, and Christopher Woodruff, "Challenges of change: An experiment promoting women to managerial roles in the bangladeshi garment sector," Technical Report, National Bureau of Economic Research 2020.

Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth, "Measuring and bounding experimenter demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

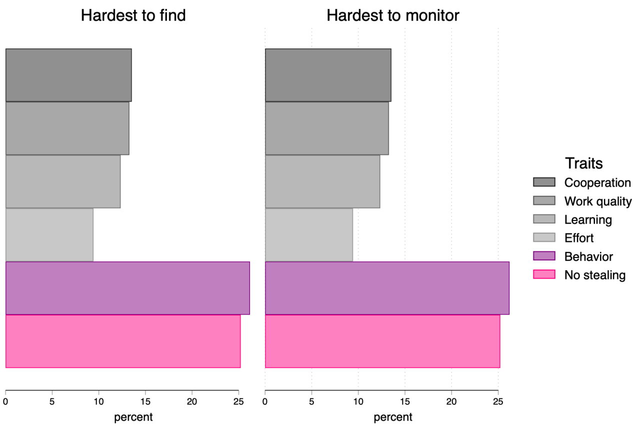
Figures

Figure 1: Valuation and Gender-Based Perceptions of Worker Traits

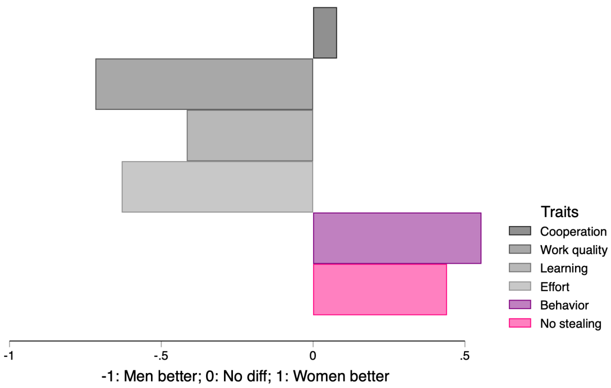
(a) “What is the single most important trait in a worker?”



(b) “Of the following workers’ traits, which ones are the...?”

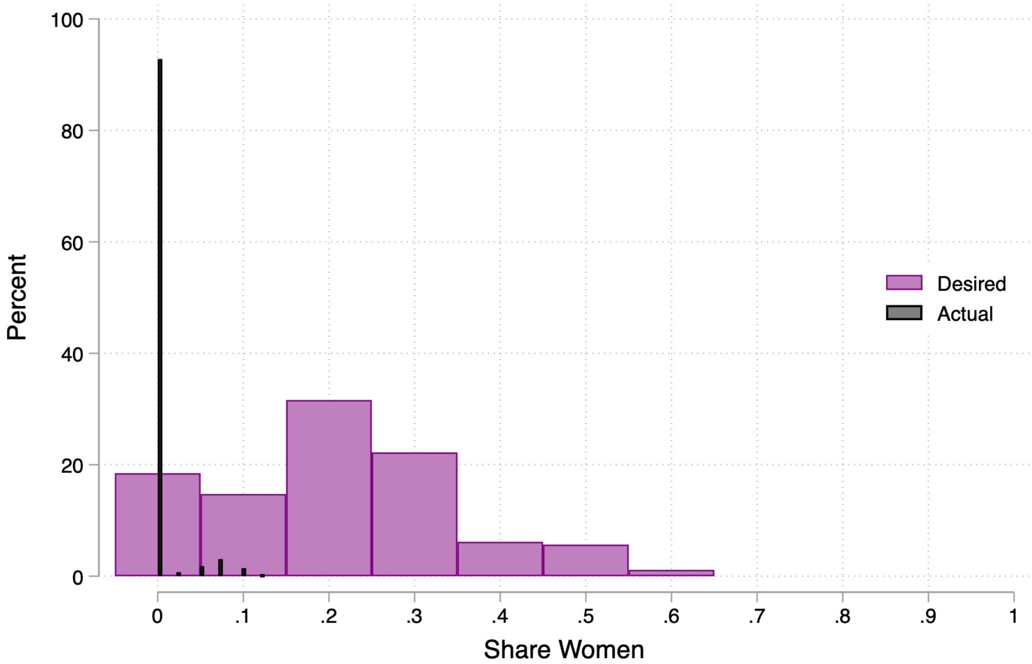


(c) “Do you think that male or female workers are better at ...?”



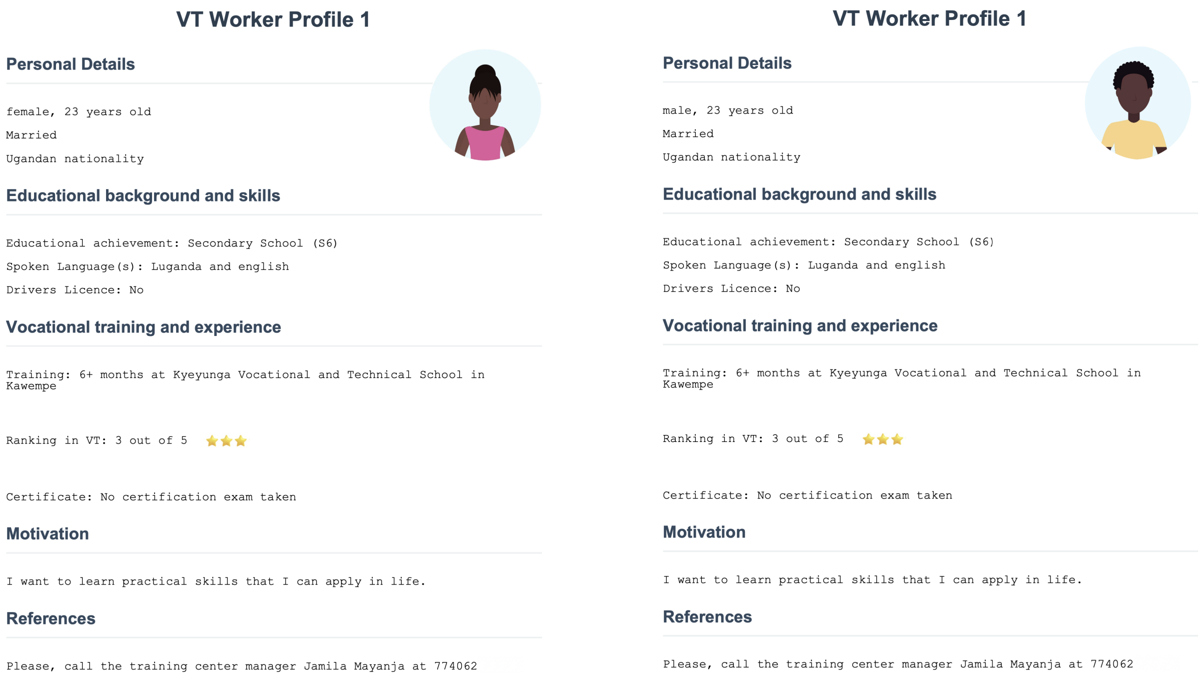
Note: Own survey data from July 2024, ($N = 617$). The question in Panel (a) was elicited through an open-ended format and re-coded. The options in Panel (b) and (c) were subsequently derived from pilot responses to the first question.

Figure 2: Managers' Desired and Actual Gender Mix Among Workers in Firm



Note: Data are based on our survey conducted in July 2024 ($N = 617$). "Desired share of women" refers to responses to the question, "What is the best gender composition of workers in your firm?" on a scale from 0 men out 10 (all women) to 10 men out of 10 (all men). "Actual share of women" is calculated as the number of women currently employed at the firm divided by the total number of workers present at the firm on the day of the interview.

Figure 3: CV Examples



VT Worker Profile 1

Personal Details

female, 23 years old
Married
Ugandan nationality

Educational background and skills

Educational achievement: Secondary School (S6)
Spoken Language(s): Luganda and english
Drivers Licence: No

Vocational training and experience

Training: 6+ months at Kyeyunga Vocational and Technical School in Kawempe

Ranking in VT: 3 out of 5

Certificate: No certification exam taken

Motivation

I want to learn practical skills that I can apply in life.

References

Please, call the training center manager Jamila Mayanja at 774062

VT Worker Profile 1

Personal Details

male, 23 years old
Married
Ugandan nationality

Educational background and skills

Educational achievement: Secondary School (S6)
Spoken Language(s): Luganda and english
Drivers Licence: No

Vocational training and experience

Training: 6+ months at Kyeyunga Vocational and Technical School in Kawempe

Ranking in VT: 3 out of 5

Certificate: No certification exam taken

Motivation

I want to learn practical skills that I can apply in life.

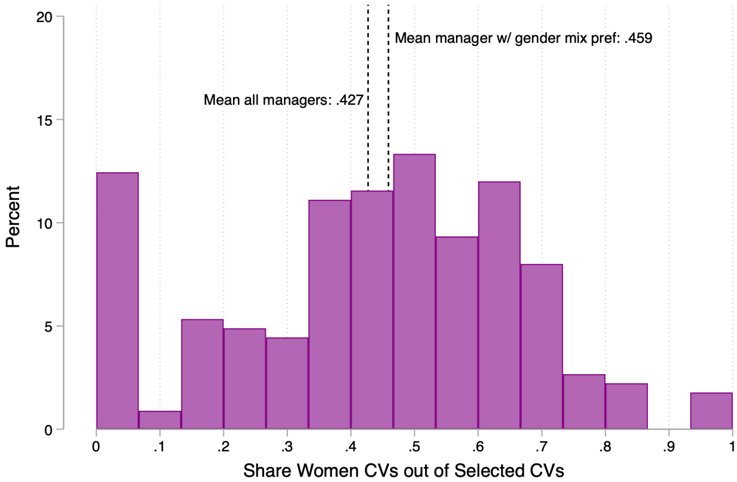
References

Please, call the training center manager Jamila Mayanja at 774062

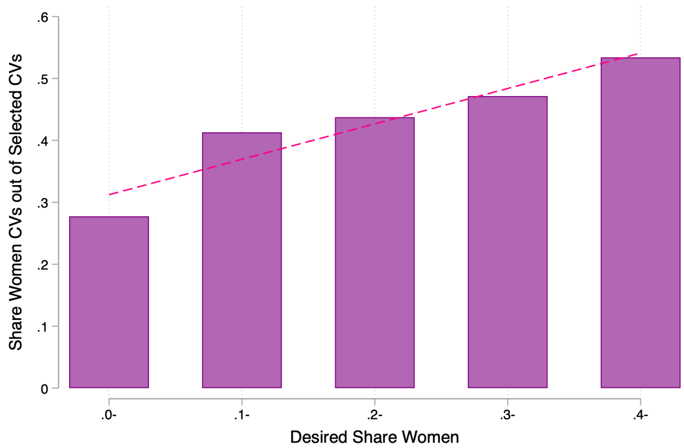
Note: Examples of CVs in both female and male versions. The content of the base profile is randomized from real applicants data. Each CV also varied in ability, indicated by a GPA score out of 5 and represented with five stars. In total, there are 24 CVs, with four variations for each base profile.

Figure 4: Proportion of Female CVs Selected by Managers

(a) Distribution



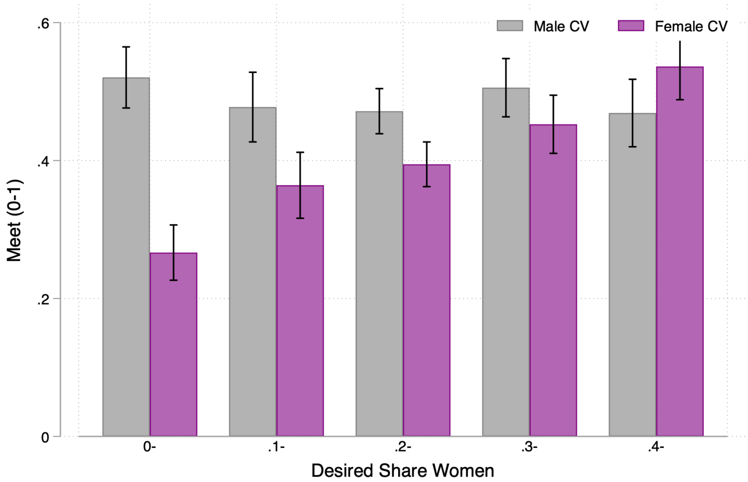
(b) Correlation with Diversity Preferences



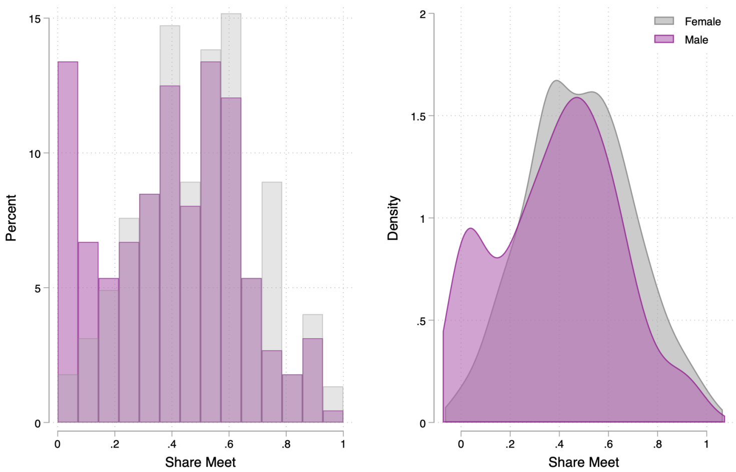
Note: The experimental data, collected in July 2024, comprises 2,396 hiring decisions made by 225 managers in the control group. Contiguous blocks of eight decisions with no variation in hiring outcomes were excluded from the sample as per the pre-registered protocol. The coefficient of the simple regression of the share of selected female CVs on the desired share women at manager level is 0.058 (p-value 0.000).

Figure 5: Gender Gap in Hiring

(a) Average Treatment Effect (Female)



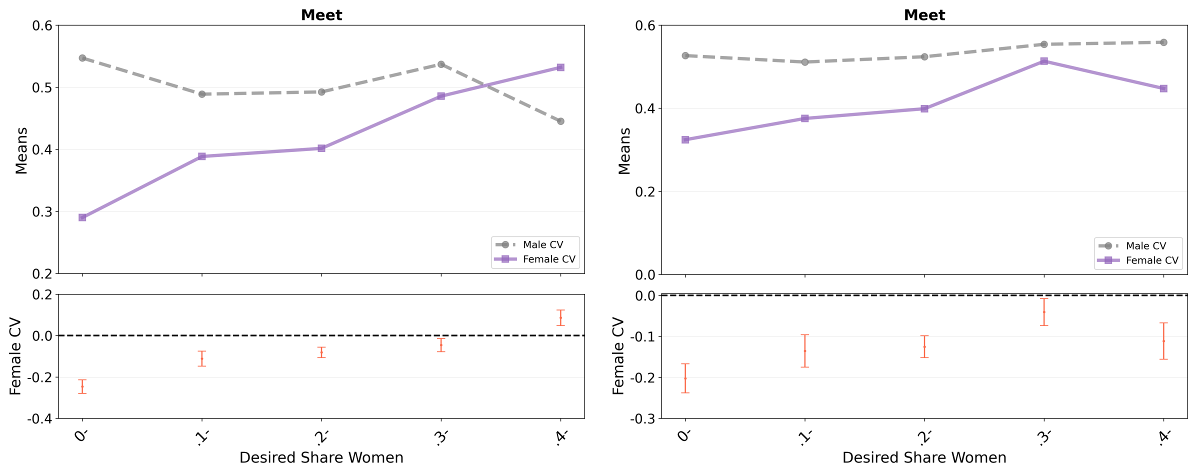
(b) Distribution



Note: The experimental data, collected in July 2024, comprises 2,396 hiring decisions made by 225 managers in the control group. Panel (a) illustrates the average differences in hiring outcomes by candidate CV gender for requests to meet (Meet). Offer was elicited on a 0-10 scale, and rescaled for comparability. Panels (b) depict the distribution of hiring for male and female CVs, respectively.

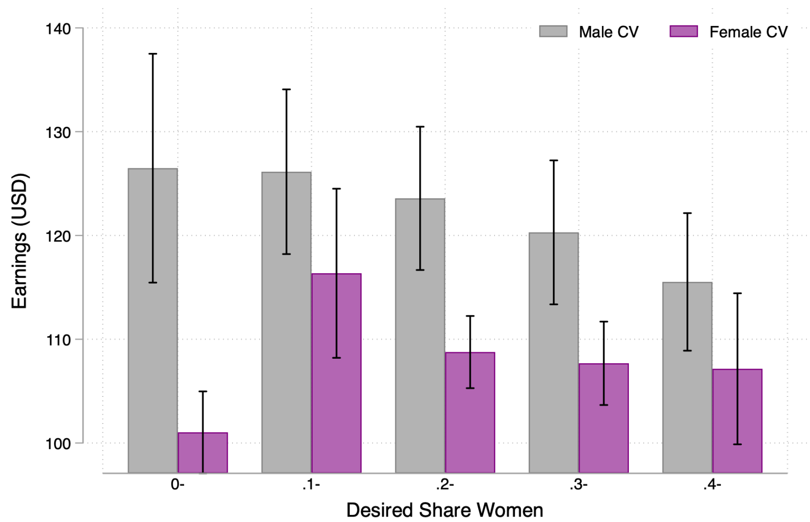
Figure 6: Gender Gap in Hiring by Diversity Preferences and Monitoring Treatment

(a) Pure Control (PC) (b) Monitoring Trainee (MT)



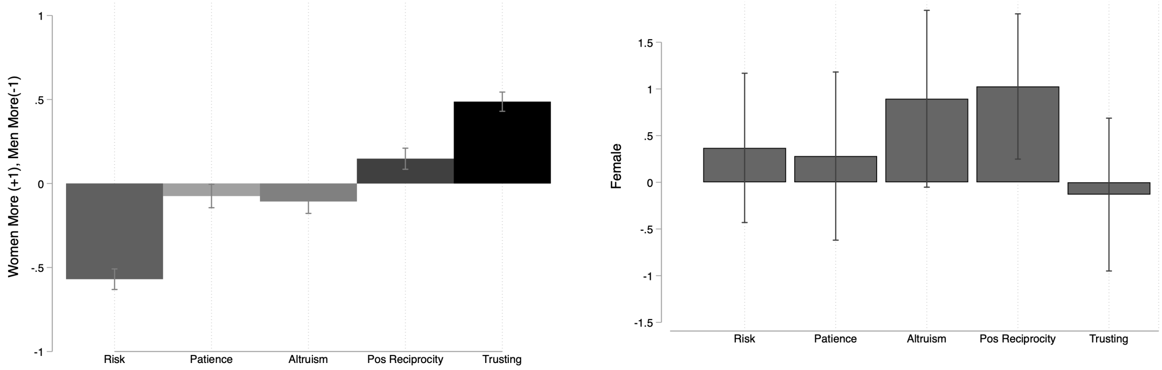
Note: Experimental data collected in July 2024. The figures plot heterogeneity in the gender gap by preferences for diversity, by treatment arm. In each figure, the left panel presents the average of the dependent variable by profile gender, while the right panel displays heterogeneous treatment effect of a CV being assigned to a female candidate, estimate with simple OLS. Confidence intervals are at the 95% level.

Figure 7: Beliefs About Labor Market Outside Options



Note: Experimental data from July 2024. Predicted candidate monthly earnings a year from now by preferences for diversity in the workplace (N= 5,390).

Figure 8: Gender Differences in Social Preferences



(a) Managers: Beliefs

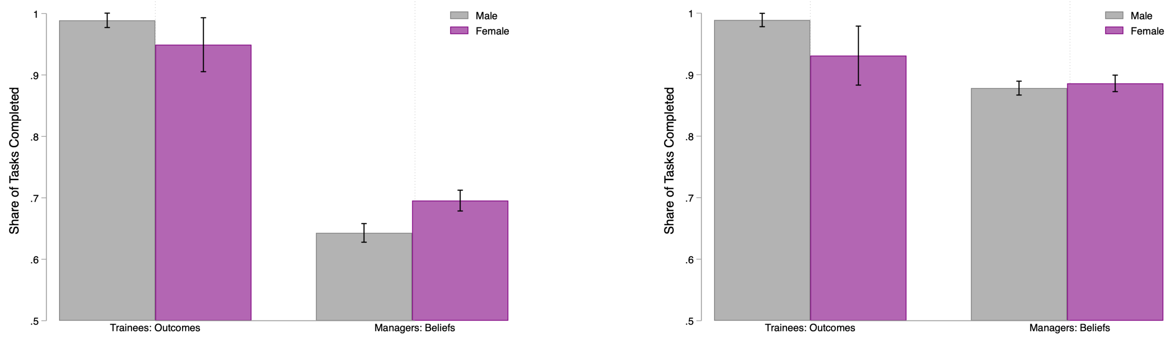
(b) Trainees: Outcomes

Note: Survey data from July 2024. We ask managers to predict the share of male and female trainees would agree with each statement. Altruism: "How willing are you to give to good causes without expecting anything in return?"; Risk-taking: "How willing or unwilling are you to take risks?"; Patience: "How willing are you to give up something beneficial for you today to benefit more from that in the future?"; Positive reciprocity "When someone does me a favor, I am willing to return it."; negative reciprocity "How willing are you to punish someone who treats you unfairly, even if there may be costs for you?"; and Trust "I assume that people have only the best intentions."

Figure 9: Gender Differences in Effort and Response to Supervision

(a) Unsupervised

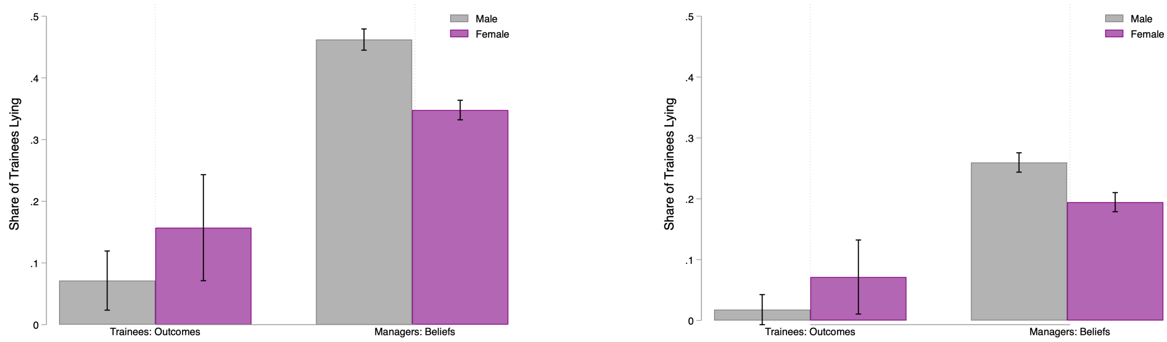
(b) Supervised



Note: Managers were asked to predict the number of tasks that male and female trainees would complete under supervised and unsupervised conditions, serving as a measure of effort. The task was designed to incentivize workers to misreport the number of tasks they had actually completed. Thus, we ask managers to estimate the extent of task misreporting under both conditions, serving as a measure of lying or misbehavior. Managers were asked to predict the number of tasks that male and female trainees would misreport, when incentivized to do so (with and without supervision) serving as a measure of lying or misbehavior.

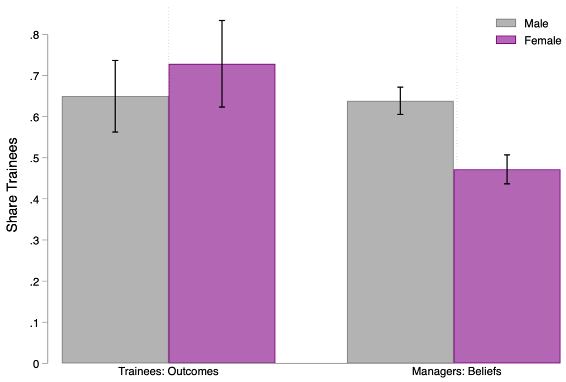
Figure 10: Gender Differences in Lying and Response to Supervision

(a) Unsupervised (b) Supervised



Note: Managers were asked to predict the number of tasks that male and female trainees would misreport, when incentivized to do so (with and without supervision) serving as a measure of lying or misbehavior.

Figure 11: Gender Differences in Technical Skills



Note: The data is from our manager and trainee survey in July 2024. Trainee data: share of trainees stating to be able to perform a certain technical task. Managers data: beliefs about share of trainees that can perform that task, by trainee gender. The three questions are sector specific. For mechanics: “Can you perform an oil change on a vehicle?”; welders: “What is the best practice for ensuring a smooth finish on a wooden surface?”; carpenters: “What is the best practice for ensuring a smooth finish on a wooden surface?”

Tables

Table 1: Descriptive Statistics

	Mean	SD	Min	Median	Max	N
Firm						
motor-mechanics (%)	0.51	0.50	0.00	1.00	1.00	617
carpentry (%)	0.31	0.46	0.00	0.00	1.00	617
welding (%)	0.00	0.00	0.00	0.00	0.00	617
firm age (years)	11.11	8.01	1.00	9.00	47.00	617
revenue monthly (usd)	611	830	0.00	314	5,555	595
capital monthly (usd)	1,474	3,983	0.00	296	37,037	597
any female worker (%)	0.16	0.37	0.00	0.00	1.00	617
family/friends hires (%)	0.41	0.49	0.00	0.00	1.00	617
n employees	10.70	10.77	1.00	7.00	70.00	617
any trainee in firm (%)	0.74	0.44	0.00	1.00	1.00	617
n clients monthly	9.41	11.48	2.00	6.00	180.00	617
recurrent customers (%)	0.51	0.26	0.00	0.57	1.00	617
Workers						
workers labor supply (h/day)	10.11	1.38	6.00	10.00	14.00	616
workers labor supply (day/week)	6.36	0.51	5.00	6.00	7.00	617
training duration (months)	12.31	7.50	0.00	12.00	36.00	617
risk for women workplace (%)	0.29	0.46	0.00	0.00	1.00	617
avg desired female workforce (%)	0.20	0.14	0.00	0.20	0.60	617
female effort higher (%)	0.57	0.50	0.00	1.00	1.00	617
Managers						
male (%)	0.96	0.19	0.00	1.00	1.00	617
owner (%)	0.85	0.35	0.00	1.00	1.00	617
ugandan nationality (%)	6.00	0.00	6.00	6.00	6.00	617
married (%)	0.73	0.44	0.00	1.00	1.00	617
any child	0.68	0.47	0.00	1.00	1.00	617
secondary educ + (%)	0.62	0.49	0.00	1.00	1.00	617
any vocational training (%)	0.33	0.47	0.00	0.00	1.00	617
respondent age	38.11	10.11	19.00	36.00	89.00	614
experience (years)	16.11	9.50	1.00	15.00	52.00	617
workdays/week	6.49	0.61	1.00	7.00	7.00	617
work hours/week	10.39	1.63	6.00	10.00	16.00	617
income 200+/month (USD)	0.55	0.50	0.00	1.00	1.00	617
Managers' Beliefs						
women more altruistic (%)	0.36	0.48	0.00	0.00	1.00	617
women more risk-taker (%)	0.18	0.38	0.00	0.00	1.00	617
women more patient (%)	0.35	0.48	0.00	0.00	1.00	617

Continued on next page

Table 1 continued from previous page

	Mean	SD	Min	Median	Max	N
women more reciprocal (%)	0.40	0.49	0.00	0.00	1.00	617
women more trustworthy (%)	0.62	0.48	0.00	1.00	1.00	617
women accepted at workplace (%)	0.54	0.50	0.00	1.00	1.00	617
women free to work (%)	0.59	0.49	0.00	1.00	1.00	617
prek child suffer if mother works (%)	0.39	0.49	0.00	0.00	1.00	617
Managers: Hiring & Monitoring						
hard to retain good workers (%)	0.33	0.47	0.00	0.00	1.00	617
screen trust hard (%)	0.64	0.48	0.00	1.00	1.00	617
screen ability hard (%)	0.32	0.47	0.00	0.00	1.00	617
monitoring constrained (h)	2.19	1.62	0.00	2.00	10.00	617
any misbehavior at firm (%)	0.85	0.35	0.00	1.00	1.00	617
manager monitoring (h/day)	3.86	2.30	0.00	4.00	12.00	617

Note: The table presents summary statistics for the study sample, with monetary values expressed in USD. It provides the mean, standard deviation, minimum, median, maximum, and sample size for a set of 43 baseline variables. These variables include information about firms, workers, and managers, as well as managers' backgrounds and their beliefs regarding gender, hiring, and monitoring.

Table 2: Balance Table

	PC (1)	MT (2)	MW (3)	(1)-(2)	(1)-(3)	(2)-(3)	N
Firm							
motor-mechanics (%)	0.547	0.508	0.485	0.040	0.06	0.02	618
	(0.50)	(0.50)	(0.50)	(0.05)	(0.05)	(0.05)	
carpentry (%)	0.280	0.305	0.342	-0.020	-0.06	-0.04	618
	(0.45)	(0.46)	(0.48)	(0.04)	(0.05)	(0.05)	
welding (%)	0.173	0.188	0.173	-0.010	0.00	0.01	618
	(0.38)	(0.39)	(0.38)	(0.04)	(0.04)	(0.04)	
firm age (years)	10.733	11.944	10.673	-1.210	0.06	1.27	618
	(8.21)	(8.60)	(7.08)	(0.82)	(0.75)	(0.79)	
revenue monthly (usd)	541	605	702	-63	-161	-97	596
	(711)	(780)	(990)	(74)	(86)*	(91)	
capital invest monthly (usd)	1827	1552	980	274	846	572	598
	(5202)	(3966)	(1663)	(454)	(372)**	(311)*	
any female worker (%)	0.164	0.162	0.163	0.000	0.00	0.00	618
	(0.37)	(0.37)	(0.37)	(0.04)	(0.04)	(0.04)	
hires close network (%)	0.360	0.452	0.418	-0.090	-0.06	0.03	618
	(0.48)	(0.50)	(0.49)	(0.05)*	(0.05)	(0.05)	
n employees	11.156	12.604	11.449	-1.450	-0.29	1.16	618
	(10.06)	(11.26)	(11.02)	(1.05)	(1.03)	(1.12)	
n clients monthly	9.409	10.208	8.617	-0.800	0.79	1.59	618
	(11.20)	(13.79)	(8.96)	(1.23)	(0.98)	(1.17)	
recurrent customers (%)	0.506	0.514	0.514	-0.010	-0.01	0.00	618
	(0.27)	(0.25)	(0.26)	(0.02)	(0.03)	(0.03)	
Workforce							
workers labor supply (h/day)	10.191	10.213	9.933	-0.020	0.26	0.28	617
	(1.34)	(1.43)	(1.36)	(0.14)	(0.13)*	(0.14)**	
workers labor supply (day/week)	6.342	6.350	6.383	-0.010	-0.04	-0.03	618
	(0.52)	(0.49)	(0.52)	(0.05)	(0.05)	(0.05)	
training duration (months)	12.071	12.660	12.173	-0.590	-0.10	0.49	618
	(7.61)	(7.77)	(7.12)	(0.75)	(0.72)	(0.75)	
any trainee (%)	0.716	0.777	0.730	-0.060	-0.01	0.05	618
	(0.45)	(0.42)	(0.45)	(0.04)	(0.04)	(0.04)	
risk for women workplace (%)	0.231	0.365	0.291	-0.130	-0.06	0.07	618
	(0.42)	(0.48)	(0.46)	(0.04)***	(0.04)	(0.05)	
avg desired female workforce (%)	0.210	0.203	0.199	0.010	0.01	0.00	618
	(0.15)	(0.15)	(0.13)	(0.01)	(0.01)	(0.01)	
female effort higher (%)	0.564	0.569	0.566	0.000	0.00	0.00	618
	(0.50)	(0.50)	(0.50)	(0.05)	(0.05)	(0.05)	
Managers							
male (%)	0.956	0.959	0.980	0.000	-0.02	-0.02	618
	(0.21)	(0.20)	(0.14)	(0.02)	(0.02)	(0.02)	
owner (%)	0.853	0.863	0.842	-0.010	0.01	0.02	618
	(0.35)	(0.34)	(0.37)	(0.03)	(0.04)	(0.04)	
ugandan nationality (%)	6.000	6.000	6.000	0.000	0.00	0.00	618
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
married (%)	0.698	0.756	0.750	-0.060	-0.05	0.01	618
	(0.46)	(0.43)	(0.43)	(0.04)	(0.04)	(0.04)	

Continued on next page

Table 2 Continued from previous page

	PC (1)	MT (2)	MW (3)	(1)-(2)	(1)-(3)	(2)-(3)	N
any child	0.618	0.629	0.663	-0.010	-0.05	-0.03	618
	(0.49)	(0.48)	(0.47)	(0.05)	(0.05)	(0.05)	
secondary educ + (%)	0.591	0.619	0.663	-0.030	-0.07	-0.04	618
	(0.49)	(0.49)	(0.47)	(0.05)	(0.05)	(0.05)	
any vocational training (%)	0.289	0.345	0.367	-0.060	-0.08	-0.02	618
	(0.45)	(0.48)	(0.48)	(0.05)	(0.05)*	(0.05)	
respondent age	37.924	37.566	38.836	0.360	-0.91	-1.27	615
	(9.87)	(9.55)	(10.91)	(0.95)	(1.02)	(1.04)	
experience (years)	15.578	16.426	16.347	-0.850	-0.77	0.08	618
	(9.20)	(9.32)	(10.03)	(0.90)	(0.94)	(0.98)	
workdays/week	6.489	6.503	6.485	-0.010	0.00	0.02	618
	(0.68)	(0.59)	(0.53)	(0.06)	(0.06)	(0.06)	
work hours/week	10.547	10.376	10.219	0.170	0.33	0.16	618
	(1.56)	(1.70)	(1.62)	(0.16)	(0.16)**	(0.17)	
income 200+/month (USD)	0.529	0.569	0.571	-0.040	-0.04	0.00	618
	(0.50)	(0.50)	(0.50)	(0.05)	(0.05)	(0.05)	
Managers's Beliefs							
women accepted at workplace (%)	0.547	0.543	0.531	0.000	0.02	0.01	618
	(0.50)	(0.50)	(0.50)	(0.05)	(0.05)	(0.05)	
women free to work (%)	0.551	0.599	0.633	-0.050	-0.08	-0.03	618
	(0.50)	(0.49)	(0.48)	(0.05)	(0.05)*	(0.05)	
prek child suffer (%)	0.404	0.381	0.372	0.020	0.03	0.01	618
	(0.49)	(0.49)	(0.48)	(0.05)	(0.05)	(0.05)	
women more altruistic (%)	0.360	0.355	0.362	0.000	0.00	-0.01	618
	(0.48)	(0.48)	(0.48)	(0.05)	(0.05)	(0.05)	
women more risk-taker (%)	0.173	0.183	0.179	-0.010	-0.01	0.00	618
	(0.38)	(0.39)	(0.38)	(0.04)	(0.04)	(0.04)	
women more patient (%)	0.356	0.365	0.337	-0.010	0.02	0.03	618
	(0.48)	(0.48)	(0.47)	(0.05)	(0.05)	(0.05)	
women more reciprocal (%)	0.427	0.406	0.362	0.020	0.06	0.04	618
	(0.50)	(0.49)	(0.48)	(0.05)	(0.05)	(0.05)	
women more trustworthy (%)	0.618	0.609	0.648	0.010	-0.03	-0.04	618
	(0.49)	(0.49)	(0.48)	(0.05)	(0.05)	(0.05)	
Managers: Hiring & Monitoring							
hard to retain good workers (%)	0.320	0.330	0.327	-0.010	-0.01	0.00	618
	(0.47)	(0.47)	(0.47)	(0.05)	(0.05)	(0.05)	
screen trust hard (%)	0.671	0.589	0.658	0.080	0.01	-0.07	618
	(0.47)	(0.49)	(0.48)	(0.05)*	(0.05)	(0.05)	
screen ability hard (%)	0.369	0.294	0.306	0.070	0.06	-0.01	618
	(0.48)	(0.46)	(0.46)	(0.05)	(0.05)	(0.05)	
monitoring constraint (h)	2.116	2.340	2.112	-0.220	0.00	0.23	618
	(1.64)	(1.63)	(1.59)	(0.16)	(0.16)	(0.16)	
any misbehavior at firm (%)	0.858	0.868	0.832	-0.010	0.03	0.04	618
	(0.35)	(0.34)	(0.38)	(0.03)	(0.04)	(0.04)	
manager monitoring (h/day)	4.054	3.563	3.939	0.490	0.11	-0.38	617
	(2.53)	(1.88)	(2.39)	(0.22)**	(0.24)	(0.22)*	

Note: The table presents a comparison of baseline characteristics of respondents by monitoring condition: Pure Control (PC), Monitoring Trainees (MT), Monitoring Workplace (MW). The differences columns are generated by a regression of each outcome on a treatment dummy with robust standard errors and strata fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The differences reports all pair-wise comparison across treatment conditions. The last column reports the sample size of the regression used to run the difference-in-means tests.

Table 3: Experimental Results – OLS Gender in Pure-Control Group

	Meet		
	(1)	(2)	(3)
Female	-0.0864***	-0.0860***	-0.0808***
	[0.0220]	[0.0209]	[0.0252]
Gender diversity (std)		-0.00947	
		[0.0157]	
Female × Gender diversity (std)		0.0960***	
		[0.0239]	
High-GPA			0.119***
			[0.0229]
Female × High-GPA			-0.0136
			[0.0272]
Observations	4576	4576	4576
R-squared	0.135	0.148	0.148
Control mean	0.504	0.492	0.415

Note: This table reports the treatment effects from estimation regression models 1, 2, and 3. *female* is a binary variable indicating whether the CV corresponds to a woman. *Gender diversity (std)* is the share of female workers in the preferred workforce composition of the manager, standardized with respect to control mean and variance. *High-GPA* is a dummy indicating whether the CV has a top GPA score (5/5) within the vocational training program. The sample are evaluations by managers in the Pure-Control group (PC). We report the primary outcome variable: interest of manager to meet (*Meet*). Profile and strata fixed effects are included but not reported in the table. Standard errors are clustered by respondent and CV profile. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Experimental Results – OLS Gender and Monitoring

	Meet					
	between-subject			within-subject		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0866***	-0.0496**	-0.0351	-0.0852***	-0.0448**	-0.0373*
	[0.0218]	[0.0204]	[0.0218]	[0.0255]	[0.0216]	[0.0194]
Monitoring trainee	0.0441**	0.0466**	0.0502**	0.0522***	0.0673***	0.0783***
	[0.0196]	[0.0204]	[0.0230]	[0.0176]	[0.0173]	[0.0183]
Monitoring workplace	0.0319	0.0233	0.0231	0.00655	0.0185	0.0249
	[0.0201]	[0.0211]	[0.0242]	[0.0198]	[0.0202]	[0.0224]
Female ×	-0.0338	-0.0524*	-0.0569*	-0.0762**	-0.0887***	-0.0956***
Monitoring trainee	[0.0299]	[0.0291]	[0.0313]	[0.0290]	[0.0267]	[0.0267]
Female ×	-0.00575	-0.0151	-0.0185	0.0352	0.0303	0.0320
Monitoring workplace	[0.0318]	[0.0301]	[0.0335]	[0.0349]	[0.0333]	[0.0345]
p-value MT vs. MW	0.3649	0.2198	0.2576	0.0029	0.0010	0.0011
Observations	12544	10364	8431	9613	8211	7239
R-squared	0.122	0.125	0.119	0.215	0.206	0.203
control mean	0.504	0.495	0.495	0.491	0.488	0.526
Preferred share female workers	any	not all men	$>$ p50	any	not all men	$>$ p50

Note: This table reports the treatment effects from estimation regression models 4 and 5. *female* is a binary variable indicating whether the CV corresponds to a woman. *monitoring trainee* and *monitoring workplace* are dummies indicating whether the respondent was assigned to MT group or MW group. The sample are evaluations by managers who passed the attention checks. The reported outcome is the interest of manager to meet a similar candidate(*Meet*). Column (1) reports the estimates of the regression model for the across-subject design 4 on the full sample. Columns (2) and (3) present the estimates of the same model 4, applied to two sub-samples: managers with gender diversity preferences excluding those with all-men preferences and managers with a preferred mixed-gender workforce above the median. Profile and strata fixed effects are included but not reported in the table. Column (4) reports the estimates of the regression model 5 for the within-subject design. Column (5) and (6) reports the estimates of the same sub-samples of columns (2) and (3) but for equation 5. Profile and strata fixed effects are included but not reported in the table. Standard errors are clustered by respondent and CV profile in both models. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.