Hidden Discrimination in Frictional Labor Markets

Elisa Macchi* and Claude Raisaro†

Version: September 8, 2025.

Abstract

We study how stereotypes and market frictions shape women's entry into male-dominated sectors and the resulting gender gap in hiring. In a field experiment with 921 employers in three male-dominated fields in Uganda, we offered opportunities to hire male and female trainees under business-as-usual conditions and two randomly assigned monitoring regimes. Employers face strong constraints in monitoring worker behavior and view women as more trustworthy. In business-as-usual, we observe a moderate gender gap in hiring, which narrows significantly among the employers with stated preferences for hiring women. When provided with audits to monitor and prevent worker misbehavior, however, the gender gap widens by 63%, with the sharpest increase among those stating the strongest pro-workforce-gender-mix preferences. By contrast, audits targeted at monitoring workers' safety — the biggest worry in hiring women — bring the observable gender gap to zero. The uncovered hiring differences cannot be attributed to ability gaps, and while about half can be explained by safety concerns, the remainder is consistent with bias. Market frictions can both mask and expose bias, and so easing them may unintentionally increase discrimination.

*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, Peter Hull, Rocco Macchiavello, Rohini Pande, Bobby Paksad-Hurson, Michael Peters, Jon Roth, Frank Schilbach, David Yanagizawa-Drott, Roberto Weber and workshop participants at BREAD, Brown, CEPR, NEUDC, MIT Behavioral Lunch, and Zurich Behavioral Lab Meeting 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), and pre-registered at the AER Registry (AEARCTR-0013698).

1 Introduction

Workers differ not only in skills but also in prosocial behavior and other traits, many of which carry gender stereotypes and become valuable in frictional labor markets. Suppose women are seen as stronger in traits that matter under such constraints. This perception may shift hiring patterns, such as easing entry in male-dominated fields, while hiding underlying bias. For example, WWII labor scarcity brought many women into the labor force, including into analytical fields such as cryptanalysis, where perceived strengths in precision, patience, and pattern recognition were valued. Yet if market conditions change and these traits lose value, gender gaps may quickly reemerge.[1]

We study whether, in labor market like Uganda (Breza and Kaur, 2025), frictions help women to enter male-dominated sectors while masking underlying discrimination. We focus on three typically "male" occupations: mechanics, welding, and carpentry. Female labor force participation is relatively high – 64.9% in Sub-Saharan Africa and 76.5% in Uganda (2025, ILO) – yet gender segregation remains stark, with most occupations nearly single-gendered. This segregation carries implications for poverty reduction (Gassier et al., 2022) and for the efficient allocation of talent (Hsieh et al., 2019; Ashraf et al., 2024).

We design a field experiment to measure how frictions affect demand for female workers, building on two novel facts. First, 84.9% of employers in the three heavily male-dominated sectors we study report a preference for hiring more women, citing low supply as the main barrier.[2] Second, facing the typical monitoring constraints of low-income settings, employers identify "good behavior, honesty, and not stealing" as the most valued yet

[1] "It was generally believed that women were good at tedious work – and, as code-breaker Ann Caracristi recalled, the early stages of cryptanalysis were very tedious," (Hammond, 2017). After World War II, women's labor force participation reverted to prewar levels once labor market conditions normalized. While causal identification is difficult, the literature attributes most of this reversion to declining demand rather than to female labor supply (Goldin, 1991; Rose, 2018). The stereotype that sees men as geniuses and women as hardworking persists in today's academia (Eberhardt et al., 2023).

[2] We find no evidence of concealed preferences for an all-male workforce, as discussed in Section 2.

hardest-to-find traits and systematically rate women as more trustworthy workers.[3] Thus, the main hypothesis we test is whether reducing supervision costs widens the hiring gender gap. We also examine the contrasting case, where frictions make stereotypically female traits more costly, as one in three employers reports that worker safety and harassment are the main worry when thinking of hiring female workers.

To measure demand for female workers and the hiring gender gap, we leverage an incentivized ratings design (Kessler et al., 2019) and address alleged supply constraints by presenting 921 employers with male-and female-assigned trainee profiles. Profiles are created by cross-randomizing administrative data from VTI students and stratifying by gender and ability. Employers evaluate twenty-four randomly-selected hypothetical résumés, for a total of 20,725 résumé evaluations. Based on these revealed preferences, we provide tailored trainee referrals. The main outcome is the decision to meet a candidate for probationary hiring.

We randomize employers to one of three conditions, subtly varying the monitoring support provided: (i) Pure Control (PC), no audit visits; (ii) Monitoring-Behavior (MB), unannounced audits to deter theft, dishonesty, and disrespect from new workers; and (iii) Monitoring-Safety (MS), unannounced audits to ensure safety and respect of new workers.[4] The PC arm measures the gender gap under business-as-usual conditions. The first treatment (MB) tests our main hypothesis: relaxing monitoring constraints increases gender differences in probationary hiring by eroding women's trust premium. The second treatment (MS) offers robustness on the mechanism (audits may either increase the gender gap in hiring against women by limiting employers' opportunities for harassment)

[3]This is a gender stereotype which also holds in the U.S. (Exley et al., 2024).

[4]The audits mimic support often provided by Vocational Training Institutes (VTIs) – such as our main VTI partner – where staff visit firms during trainees' internships. Our design builds on this institutional practice so that the visits are familiar to employers.

and/or decrease it by relaxing women-specific monitoring constraints — if employers fear for women's safety on the workplace.[5]

Are employers actually interested in hiring women? When offered the opportunity to hire trained women on probation under business-as-usual conditions (PC), employers select an average of 47% of the trainees they evaluate, and 44.1% of all selected candidates are women. The gender gap, conditional on observables, is 10.3 percentage points and narrows significantly with stated preferences for hiring more women. There is no gender gap in the top quintile of the distribution — employers who report wanting 4 to 7 women per 10 workers. The relatively mild gap is comparable with evidence from Uganda in similar sectors (Alfonsi et al., 2020), and provides support for the view that there is demand for female workers in these male-dominated sectors in the status quo.[6]

We then test how the observed hiring gender gap varies under the different monitoring regimes. First, we look at the effect of relieving employers' monitoring constraints regarding workers' behavior and theft. As hypothesized, employers receiving support to monitor trainee behavior and prevent theft (MB) show a larger gender gap than the pure control. The gap under MB is 16.8 percentage points — a 63% increase relative to the pure control. The larger gender gap is driven by firms with stronger stated preferences for workforce gender mix. Among employers with some preference for hiring women, the gap doubles; among those above the median, it is more than three times larger. Thus, relaxing monitoring constraints reveals latent gender gaps, particularly among employers who appeared more gender-neutral at baseline.

Next, we look at the effect of providing safety audits (MS). Safety audits may influence the hiring gender gap through two opposing channels. On the one hand, they may encourage

[5]We label the MS treatment "Safety" to emphasize its gender-neutral framing around safety, harassment, and respectful treatment.

[6]For comparison, Buchmann et al. (2023) document a gender gap exceeding 40 percentage points in a male-stereotypical occupation in Bangladesh.

the hiring of women on probation by alleviating concerns about harassment from coworkers or customers. On the other hand, they may deter employers from hiring women on probation by introducing a scrutiny effect. We find support for the safety-concerns channel over the scrutiny channel. Among employers assigned to audits that monitor worker safety (but not behavior) the gender gap in probationary hiring is not statistically significant and as small as 1.7 percentage points − a 6-fold reduction relative to the business-as-usual gender gap. The results strengthen our interpretation of the impact of behavior audits (MB) on gender gaps by ruling out indirect audit effects related to increased scrutiny. They also help explain why few employers capitalize on women's trust advantage in the status quo: beyond search costs, safety concerns are a key barrier.

Our findings support the hidden-discrimination channel: the widening gender gap when monitoring constraints are lifted is consistent with a reduced need for trustworthy workers. Complementary survey and lab-in-the-field evidence also point to positive, trust-based statistical discrimination in favor of women, though these beliefs are inaccurate. Employers do not view women as more prosocial overall, but in behavioral games, they expect female workers to cheat less and perform better when unsupervised, despite no actual gender differences in behavior. In our main experiment, we estimate that at least 57% of the demand for women in the status quo can be explained by a demand for trust.

A remaining question is whether the gender gap we uncover reflects bias against women. First, while part of the hidden gender gap can be attributed to employers' safety concerns, at least 40% remains unexplained.[7] Next, we combine experimental data with survey evidence to investigate whether it can be explained by employers' beliefs about differences in trainee performance. While employers hold (inaccurate) gender-based priors about ability, we find no evidence that these beliefs drive hiring choices in a way consistent

[7]This is a conservative estimate, given that safety concerns are likely paternalistic and in turn, also represent discrimination (Buchmann et al., 2023).

with statistical discrimination. Employers respond to ability signals, such as résumé qualifications, but do so similarly for male and female candidates. We interpret these results as evidence that a significant share of the observed gender gap is driven by bias against women.[8]

Our paper contributes to three strands of literature. First, to the labor economics literature on discrimination, which models bias as arising from uniform and stable penalties attached to observable traits (e.g., Goldin and Rouse, 2000; Bertrand and Mullainathan, 2004; Kline et al., 2022; Macchi, 2023). Using experimental variation, we demonstrate that interventions that reduce market frictions — such as lowering monitoring costs or improving workplace safety — can shift the observable gender gap by interacting with perceptions of minority workers' comparative advantages.

We show that observed gaps can vary with the institutional or technological environment in which hiring decisions are made. This insight complements recent work by Bohren et al. (2022), who show that discrimination can accumulate across sequential stages of evaluation, and that standard correspondence studies may overlook these systemic effects. We show that the "system" affects measures of discrimination even within a single evaluation node. This reveals a new layer of hidden discrimination — one that is missed by both traditional and systemic discrimination measures.

Second, we contribute to the literature on female labor force participation and gender gaps in low-income countries. While many studies examine supply-side constraints (see Jayachandran, 2015 for review), evidence on demand-side factors remains limited (Gentile et al., 2023), particularly in the African context and in sectors where women are underrepresented. Also, open questions remain about the circumstances in which training targeted at jobs where women are underrepresented can be particularly effective (Heath et al.,

[8]The results are also in line with the fact that employers often cite reasons for not hiring women, such as physical strength, that they do not themselves consider important for job performance.

2024). Our study provides novel experimental evidence from employers in three typically male-dominated sectors in Sub-Saharan Africa. The experiment simulates what would happen if trained women entered employers' hiring networks, providing insight into the demand-side barriers that shape the effectiveness of training interventions.

Our evidence for hiring gender gaps against women is consistent with previous research from settings where women are relatively well represented (e.g., Brown (2023) on pay for female teachers in Pakistan; Macchiavello et al. (2020) on promotion patterns in garment factories in Bangladesh). Particularly, Buchmann et al. (2023) show that paternalistic beliefs around workplace safety can induce gatekeeping in male-dominated occupations in Bangladesh. We build on this research in two dimensions. We show that, despite biases, positive stereotypes can enable women's entry into male-dominated fields when frictions like monitoring constraints are present. We further show that simple safety audits are effective at reducing perceived safety risks on the employer side and to an increase in women's entry.

Finally, we connect to a literature on social norms (e.g., Jayachandran, 2021; Bursztyn et al., 2020; Bernhardt et al., 2018), beliefs (e.g., Feld et al., 2022; Coffman et al., 2021), and stereotypes (e.g., Bordalo et al., 2019; Carlana, 2019) and their negative impact on minorities' outcomes. We find that negative stereotypes about women's abilities can coexist with positive stereotypes about their trustworthiness and prosocial behavior, in line with Exley et al. (2024).

2 Setting, Data, and Motivating Facts

In Uganda, female labor participation in 2024 was 76.5%, a statistic larger but in line with the 64.9% sub-Saharan Africa average (International Labour Organization, 2025).

Despite such high participation, occupational gender segregation remains extremely pronounced, with most jobs single-gendered. Given the well-documented talent misallocation implications (Hsieh et al., 2019), and the evidence that women earn significantly more in male-dominated sectors than in traditionally female ones (Goldstein et al., 2019), facilitating women's entry into these fields is a key policy priority in low-income countries.

We focus on three manufacturing sectors: mechanics, welding, and carpentry. These are emerging urban working-class occupations across both formal and informal settings, and remain heavily male-dominated. We collaborate with seven major vocational training institutes (VTIs) in Kampala, Uganda, to collect data on trainees' backgrounds, job search behaviors, and implement all treatments.[9]

In addition to providing training, these VTIs routinely facilitate internships with local workshops to support entry into the labor market. Teachers typically visit trainees during internships to monitor their progress and participation. We leverage this existing practice to implement our experimental design within a well-established framework for firms.

2.1 Sample Selection and Descriptives

The primary study population for the experiment is employers from motor mechanics garages, carpentry, and welding firms in the Greater Kampala area. Our main sample consists of 921 employers, one per each firm, comprising 318 from motor mechanics garages, 300 from carpentry shops, and 303 from welding firms (see Table 1 and Appendix Table A1).

[9]Our study began in collaboration with the Smart Girls Foundation, a nonprofit organization that delivers nationally recognized vocational training programs. The organization specializes in preparing women for traditionally male-dominated sectors — such as automotive mechanics, electrical installation, and welding — while also offering training in entrepreneurship and life skills (see Appendix Figure A2). Although the program primarily targets women, it also trains a minority of male students.

We randomly selected firms in the neighborhoods of the greater Kampala area where firms in those sectors are clustered, as shown in Figure A1.[10] Respondents were classified as employers if they had discretion over hiring decisions. We selected employers: (i) working in small and medium-sized firms (fewer than 100 employees); (ii) aged 18 or older, (iii) expressing interest in hiring a trainee within nine months; (iv) proficiency in either English or Luganda, (v) with hiring discretion. In our sample, 84.9% of the respondents are business owners, while 15.1% hold managerial positions. The firms are well established and employ a mean of 8.3 workers (median 5).[11] They serve approximately 11.8 clients per day, have operated for 10.7 years, and report mean monthly profits of USD 286 on revenues of USD 718.[12]

The sectors are overwhelmingly male-dominated. 96.4% of employers in our sample are men, and only 2.26% of all workers are women (173 out of 7,658 employees). Fewer than 14% of firms employ any women at all. Previous research on samples for welding and motor-mechanics firms in Uganda reports a share of female workers around 1 to 3% (Alfonsi et al., 2020).

The organizational and physical structure of firms makes it difficult for employers to easily monitor workers. Workshops typically include indoor and large outdoor areas (see Figure A3 for an example). Due to the relatively low specialization (Bassi et al., 2023), employers typically combine technical tasks with supervision and on-the-job training of new hires. As a result, they must divide their time between production and monitoring, and nearly all face binding monitoring constraints.

[10] Within each parish, the field research assistant identified a central landmark and assigned a random direction for each field officer to follow while searching for respondents.

[11] The average size of our sample is also in line with the data from the Census of Business Establishments for Uganda (UBOS, 2011).

[12] Uganda's 2024 GDP per capita is USD 1,023.

On average, employers spend 10.4 hours at the firm each day, dedicating 2.06 hours to monitoring — making it the second most time-consuming activity after technical work, and more time-consuming than training. Over 90% report a desire to increase monitoring, with an ideal increase of approximately 30% in monitoring hours. Theft is a common concern: 83% of employers report having experienced theft by an employee.

Likely as a consequence of the monitoring constraints faced by firms (Heath, 2018), hiring practices are highly informal: about 70% of employers recruit through personal referrals, and 44% employ at least one family member or close friend. In line with Alfonsi et al. (2020), interns in our study face varied payment structures: most are unpaid, some pay for training, and others are paid. The median probationary worker receives no remuneration; 47% earn a positive wage, 64% pay for training, and in general only 15. 9% receive net positive pay.

Our second population of interest consists of trainees enrolled in automotive mechanics, carpentry, or welding programs at partner VTIs. Through our partnership with the VTIs, we collect administrative data and survey 182 trainees. 33% of the trainees specialize in automotive mechanics, 19.8% in carpentry, and 21.2% in welding.[13] Trainees, 21.6 years of age on average, are enrolled in programs lasting between 6 and 24 months, with those longer than one year accredited by the Uganda Directorate of Industrial Training (DIT).

Women make up 37% of our sample; the share falls to 18.8% if we exclude trainees from *Girls With Tools* VTI — our largest provider of female candidates. This represents a reasonable upward trend compared to the 11% share of female vocational training applicants in welding and motor-mechanics documented by Alfonsi et al. (2020). Nationally, in 2023, women were 8.8% of trainees who completed a certification in our target sectors under the Uganda Directorate of Industrial Training (Uganda, 2023). In comparison, our sample

[13] An additional 16.5% are enrolled in broader fields such as electrical work but report job aspirations aligned with our target trades.

also includes shorter, non-DIT training programs, and focuses on Kampala metropolitan area.

2.2 Preferences and Beliefs about Skills, Trust, and Gender

Stated Preferences for Workforce Gender-Mix. We document employers' stated preferences for gender composition in the workplace. When asked "What is the best gender composition of the workers in your firm?" — on a scale from 0 to 10 men out of 10 — 84.9% of employers report preferring at least one woman on a ten-person team, and 70.5% prefer two or more (Figure 1, panel (a)). 14.6% of employers state a preference for a roughly gender-balanced team of 4 to 6 women out of 10.[14] Given that 86.3% of firms currently employ no women, 92.3% stating a preference for a gender-mixed workforce suggests an unmet demand for female workers.

Perceived Barriers to Hiring Female Workers. Employers estimate that fewer than 5% of prospective workers in the applicant pool are women − on average, 0.5 out of 10. Various frictions may prevent employers from finding female workers, as we investigate in Figure 1, panel (b). When asked why they do not currently hire any women, over half of employers cite supply-side constraints. These frictions are consistent with employers' reliance on personal networks − friends, family, clients, or current employees − for over 80% of their hires. In male-dominated sectors, such network-based recruitment has been shown to reinforce gender imbalances in favor of men (Beaman et al., 2018; Alfonsi and De Souza Ferreira, 2024).

[14]These preferences are self-reported and not incentive-compatible, which motivates our experiment to measure revealed preferences. Yet it is reassuring to see no bunching at focal points like 1 or 5, which suggests there is no systematic overreporting of female or measurement error.

When asked about potential concerns if they were to hire a woman, the most frequent worry is safety and harassment, mentioned by over 40% of employers (Figure 2, panel (b)).[15] This suggests that firm-level constraints influence both women's hiring and the effectiveness of training and matching interventions, consistent with Bandiera et al. (2022).

Negative stereotypes and behavioral factors are also likely at play, though they seem most relevant for the minority of employers (about 15%) who explicitly state they do not want to hire women. More generally, employers often cite physical strength (24.55%) and effort (20.16%) as worries when hiring women, yet neither ranks among their top hiring criteria, suggesting ex-post rationalization consistent with a somewhat widespread belief that these are not "jobs for women."

Gendered Perception of Workers' Skills. Employers hold gendered perceptions of both technical and non-technical skills. We asked employers to compare men and women on the traits that they value the most, their responses reflect strong gender stereotypes: men are seen as stronger in technical domains (e.g., work quality, interest and learning ability) while women are perceived as superior in trust-related traits, including good behavior, described as trustworthiness, honesty, and no stealing (Figure 1, panel (a)).

Importantly, good behavior is also the trait employers most often cite as the single most important trait in workers at the firm as well as the hardest to find and monitor (Appendix Figure A5).[16] These answers point to a widespread demand for trustworthy workers among employers in our sample — consistent with the significant monitoring constraints reported by employers. Thus, employers perceive women as having a comparative advantage in precisely the areas where frictions are greatest.

[15] Question: "If you think of hiring a woman, what do you worry about?".

[16] In contrast, while effort and hard work are the second most valued traits, only 34% of employers cite them as hard to monitor — likely because 4 out of 5 workers are on piece rates, which make effort contractible.

Beliefs about Customers and Coworkers Attitudes. Finally, we find little support for expectations of customer discrimination or coworker resistance in dampening preferences for gender mix in the workforce. In fact, customers are not mentioned as a barrier to hiring women or as a concern when considering them (Figure 1). Moreover, cooperation is not a trait highly valued in hiring decisions[17], and female workers are if anything perceived as better cooperators than men (Figure 1). Approximately 65% of employers expect "all" of their current workers to accept a female colleague, with the average expected acceptance rate of 82%.

3 Experimental Design

We design an experiment to address employers' stated primary constraint to hiring women: limited supply (Fig. 1). To test whether demand exists once this constraint is relaxed, we use an incentivized ratings design (Kessler et al., 2019): employers seeking to hire trainees are asked to evaluate hypothetical trainee profiles, in order to receive referrals of real candidates for probationary hiring at their firms.

Beyond supply, the experiment also targets two frictions related to women's entry: monitoring constraints and safety/harassment concerns. At the firm level, we randomize the type of audit support provided — behavior audits or safety audits — as technologies to relax these constraints. The use of audits is also directly in line with ILO recommendations and thus, policy relevant. In what follows, we describe the profile evaluation exercise and the treatment assignment.

[17] As shown in Appendix Figure A5, only 2.27% of employers cite "cooperation" as the most important worker trait.

3.1 Profiles: Gender and Ability Randomization

To construct the hypothetical trainee profiles, we collect administrative data on 472 enrolled trainees from seven partner VTIs in Kampala. Using information from VTI administrative records, we create 36 hypothetical profiles by cross-randomizing key attributes such as age, marital status, possession of a driver's license, language spoken, motivation for career choice, education, training background, and references.[18] Each of these 36 profiles is then used to generate four versions, yielding a total of 144 unique résumés. This is achieved by cross-randomizing two additional attributes: gender, signaled by an avatar, and ability, indicated by class rank and visualized through a star rating (three stars for average performance, five stars for top performance). Figure 3 illustrates an example of how résumés were presented to employers. Each employer is randomized to review one of the four versions of each profile.

We ask employers to evaluate a randomly assigned set of 24 hypothetical candidate profiles during their regular working hours.[19] Our primary outcome is the employer's decision to initiate a probationary hire, and we also collect a set of secondary outcomes to probe employers' beliefs and preferences in greater detail−specifically, perceptions of work quality, behavior, and expected future earnings.[20] This design allows us to evaluate differential responses by gender and ability in 2×2 design, with random variation across employers for each profile. Each employer sees one randomly-selected version for each profile.

[18]The information provided in the résumés is chosen based on initial focus groups with a separate pilot sample of 25 employers.

[19]We calibrate the number of evaluations to reflect the average monthly inflow of jobseekers (24.6) applying, as reported by pilot sample of employers during scoping activities.

[20]The primary outcome is *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/honest." *Earnings*: "What is your best guess of the monthly earnings of this worker a year from now?"

While not central to our design, the cross-randomization of candidate gender and ability allows a standard test for statistical discrimination (Bertrand and Duflo, 2017): if gender gaps in hiring choices narrow as candidate ability increases, it suggests employers rely on gender as a proxy for ability when information is limited.

Employers' choices are incentive-compatible by design. While employers are told that the profiles are hypothetical, they are also informed that, at the end of the study, referrals will be made from a pool of real candidates drawn from our partner vocational training institutes. This creates a direct link between employers' evaluations and the potential to receive actual trainees. Each respondent is carefully guided through the incentive structure prior to beginning the profile evaluation exercise, ensuring that they understand the stakes and relevance of their decisions.

3.2 Treatment Assignment: Monitoring Regimes

We randomly assign employers to one of three experimental arms: one control and two different monitoring support regimes. In the Pure-Control (PC) arm, employers make choices under business-as-usual conditions to measure demand for female workers and the gender gap in probationary hiring, given that our intervention removes supply constraints.[21] In the monitoring regimes, firms receive weekly unannounced audits from members of our team. This approach builds on standard VTI practices: staff occasionally visit firms during trainees' internships to verify attendance, especially when these are curricular.

The two monitoring regimes are as follows. Employers in the Monitoring-Behavior (MB) arm receive unannounced audit visits to oversee trainee misbehavior and theft, the main

[21] We inform these employers, as in all other groups, that the research team aims to ensure "employers and workers will have a positive experience if they are matched."

constraint they report in managing their workforce (Fig. 2, panel (a)).[22] Employers in the Monitoring-Safety (MS) arm receive unannounced visits to ensure worker safety and respectful treatment.[23] This targets the main concern about hiring women (see Fig. 1, Panel (b)), although worded in gender-neutral terms to avoid priming.[24]

We emphasize that the audit assignment is random to prevent it from being seen as a signal of worker or firm quality. This ensures that any treatment effects are not driven by employer beliefs about differences in the applicant pool across monitoring conditions. We also inform employers that trainees will be informed about the audits too, once the match is finalized. Because of this, our treatment effects capture both the direct impact of monitoring and employers' expectations about how trainees will respond to anticipated oversight. This choice reinforces the intensity of the monitoring intervention, even though it remains relatively light-touch.

We expect Behavior audits to reduce employers' monitoring constraints related to misbehavior and stealing. Notably, employers' expected audit duration (1.7 hours per day) closely matches their reported gap in monitoring capacity. In turn, we predict that the

[22]The MB script is: "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 new workers are placed during the probation period. These 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, disrespect by way of monitoring."

[23]The MS script is: "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 new workers are placed during the probation period. These workers will also be informed about the visits. Your firm may be randomly selected to receive such support visits via a lottery. The visits received do not reflect the characteristics of the firm or the workers. 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, and disrespect by means of monitoring."

[24]In the first sector (mechanics) of the data collection, we implemented a *between-subjects* design. Analysis revealed far greater heterogeneity in demand for women than anticipated; in particular, different from pilot data, around 15% of the employers stated no preferences for gender-mix in the workplace, reducing statistical power beyond our expectations. To address this, we amended the pre-registration, and extended the data collection to three sectors. To be on the safe side, for the remaining sectors we introduced a *within-subject* component: after completing 24 ratings under the main treatment condition, each employer evaluated 12 additional profiles under a second monitoring condition. The main specification remained the between-subjects design; we report the within-subjects results in the Appendix.

gender gap in probationary hiring — the selection rate for male over female workers — will widen under behavior audits (MB) relative to business-as-usual (PC). We expect this effect to be driven by a reduction in the comparative value of trustworthiness, mostly associated with women, decreases.

Safety audits may affect hiring preferences through multiple channels. First, although our focus is on women, we deliberately framed the audits as gender-neutral to avoid priming employers about gender or harassment. If safety concerns are limited to issues like machine handling, audits may boost overall probationary hiring without affecting gender gaps. Second, if employers associate safety concerns primarily with women—such as the perceived cost of maintaining a harassment-free workplace, audits may encourage hiring women and reduce the gender gap. Finally, audits could increase the perceived cost of hiring women by exposing employers to greater scrutiny, potentially deterring them from hiring women and widening the gap.

Testing whether safety audits increase the gender gap (the third channel) is crucial for isolating the effect of behavior audits (MB), as scrutiny effect may be an indirect effect of any audit. More broadly, the reduced-form effect of safety audits on gendered hiring choices is policy relevant: it speaks to whether such audits can address paternalistic discrimination, a barrier to women's entry into male-dominated fields.

4 Main Results

The main sample consists of 906 employers who, as preregistered, pass attention checks related to treatment assignment and exhibit variation in profile evaluations.[25] Table 2 shows

[25]The sampling protocol is uncorrelated with treatment assignment (Appendix Table A2).

covariate balance across experimental conditions, including employers' sociodemographic characteristics and firm-level attributes under different monitoring regimes.[26]

We also analyze heterogeneity by employers' baseline stated preferences for workforce gender mix. As shown in Figure 1, the preference distribution is heterogeneous. About 15% of employers report no preference for hiring women (the bottom quintile), while the top quintile has preferences close to a 50-50 gender mix, which in turn implies a much stronger unmet demand for female workers.

4.1 Demand For Women and Hiring Gender Gap under Business-As-Usual

We analyze employers' gender-specific hiring preferences using data from the Pure Control (PC) group, where probationary hiring decisions are made under business-as-usual conditions. We measure demand for female workers descriptively as the share of female résumés among all résumés an employer chooses to meet in order to hire on probation.

Most employers show some demand for female candidates: 89.6% of employers select at least one female trainee résumé. Among all selected candidates (47% of the total résumés shown), the average share of women is 42.2% (median 44.1%), as shown in panel (a) of Figure 4. The distribution of female trainees selected is highly heterogeneous: while 10.4% of employers select no female trainees, about 35% of employers selects more female than male applicants.

Stated preferences for workforce gender mix are positively correlated with employer's selection of female candidates in the experiment. A 10-percent increase in the share of

[26] We find mild imbalances in two baseline variables: employers' perceived risk of hiring women and prior VTI training. Results are robust to including these variables as controls.

female workers the employer states to prefer predicts a 4.2 percentage point increase in the likelihood of selecting female workers (p-value < 0.001). When turning to the relative demand of female versus male workers, employers with stated preferences for a gender balanced workforce are an all-male workforce 2.4 times more likely to select a female worker relative to a male worker.

We estimate the gender gap in probationary hiring, defined as the difference in selection rates between male and female candidates with otherwise identical résumés, using the following regression model:

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

Meet_{ijs} denotes employer j's interest in meeting a profile with characteristics of résumé ij in strata s. Table 3, panel (a), summarizes the results. Female_{ij} equals 1 if employer j is randomly assigned to the female version of profile i. The coefficient β_1 captures gender bias as the differential effect of female versus male candidates on outcome Y. δ_i and σ_s are profile and strata fixed effects. Standard errors are clustered at the employer and profile level. We estimate the regression model on the full sample and by quintile of employers' stated preferences for gender mix.

On average, as shown in column 1 of Table 3, panel (a), women are 10.3 percentage points less likely to be selected for a probationary meeting as compared to men with identical résumés. This gender gap, which is equivalent to an increase of 1.3 GPA, is significant (p-value < 0.001) but relatively mild given the extreme gender segregation of the Ugandan labor market (2.3% of female workers in our sector). For comparison, recent experimental work testing for on discrimination in male-dominated fields in Bangladesh finds a gender

gap that exceeds 40 percentage points (Buchmann et al., 2023), underscoring that our context features comparatively moderate levels of discrimination.

During the probation period, the wage structure typically combines a small daily wage paid by the employer with a contribution to the internship paid by the trainee (i.e., the probation worker). For a subsample of employers in welding and carpentry, we also collected data on the expected net wage (i.e., net of any contributions the worker would pay during the probation period) for each candidate. Specifically, we asked employers: (i) the wage they would require the trainee to pay for the probation period, and (ii) the wage they themselves would pay for the probation period. We find no evidence of gender differences in net wages, as shown in Appendix Table A6.

Importantly, we do not have reasons to believe that this finding is an artifact of our experimental design, and argue that the results possess strong external validity. Indeed, the hiring patterns by gender vary meaningfully with stated preferences for the gender composition of the workforce. As anticipated, the gender gap is widest among employers who state no preference for hiring women.[27] However, this gap narrows significantly with employers' stated preferences for a more gender-balanced workforce, as evident from Figure 4, panel (b). In the regression framework, a 10%-increase in the share of female workers in an employer's ideal workforce composition is associated with a 23.8% smaller gender gap relative to employers who report all-male ideal workforce composition (p-value < 0.001).[28] We detect no statistically significant gender gap against female candidates when focusing on employers in the top quintile, those with a stated ideal gender mix of at least 40% women.

[27]Employers with no stated preferences for hiring women are 5 times more likely to select no women candidates in the experiment. We observe 33 employers in PC who hire no women, express no preference for a gender-mixed workforce, yet select one or more female résumés in the experiment. When asked to explain their choices, over 30% of these respondents cite honesty and trustworthiness as the reason for selecting female candidates. The remaining employers state that they chose female applicants to give them a chance to learn.

[28]See Appendix Table A11, Column (1).

Our data show that most employers select both male and female trainees when given the opportunity, resulting in a surprisingly modest gender gap. Considering the extreme gender segregation in these sectors, this finding raises a key question: is this pattern the result of low bias, or is the demand for women explained by a demand for trust?

4.2 Addressing Market Frictions and Hiring Gender Gap

We examine the effect of the monitoring regimes on employers' hiring preferences, focusing on the gender gap against women as outcome. First, as a sanity check, we assess the overall effect of the monitoring regimes on labor demand, measured by the number of résumés selected for probation meetings. Since both audit treatments are designed to ease monitoring, we expect them to increase the rate of selection. Indeed, as shown in Appendix Table A7, employers are 4.1% more likely to select trainees in the Monitoring-Behavior arm (p-value = 0.090), and 6.4% more likely in the Monitoring-Safety arm (p-value = 0.012).

To identify the average treatment effect of monitoring regimes on the gender gap, we estimate the following regression model on the main sample of employers:

$$\begin{aligned} \text{Meet}_{ijs} = \beta_0 + \beta_1 \text{Female}_{ij} + \beta_2 \text{MB}_j + \beta_3 \text{MS}_j \\ + \beta_4 \text{Female}_{ij} \cdot \text{MB}_j + \beta_5 \text{Female}_{ij} \cdot \text{MS}_j + \delta_i + \sigma_s + u_{ijs} \end{aligned} \tag{2}$$

Meet_{ijs} denotes employer j's interest in meeting a profile with characteristics of résumé ij in strata s. MB_j and MS_j are dummies for being assigned to the Monitoring-Behavior arm and Monitoring-Safety arm. The Pure-Control arm is the omitted group.

We also estimate a fully interacted version of the regression model to assess treatment effect heterogeneity by employers' stated preferences for the gender composition of their

workforce. This heterogeneity analysis is important because employers with no stated preference for hiring women are five times more likely to select no female trainees in the experiment, leaving less scope to reduce the gender gap by increasing female selection.

4.2.1 Do Behavior Audits Increase the Gender Gap?

We compare the selection rate of male and female trainees between employers randomly assigned to receive Behavior audits (MB), a "technology" that reduces monitoring costs, and those in the business-as-usual condition (PC).

The raw data, plotted in Figure 5, show that male trainee selection rates increase under MB relative to the business as usual condition, while the average selection rate of female trainees declines. These effects are evidently stronger among employers with stronger stated preferences for hiring women, suggesting that reduced monitoring costs reveal latent barriers that even pro-diversity employers face when evaluating female candidates.

Using our estimate of Equation 2, employers in the Monitoring-Behavior arm were 16.4 percentage points less likely to hire a female trainee, conditional on them being otherwise identical to a male trainee (Table 4, column 1). This treatment effect (β_4) captured in Figure 6 reflects a statistically significant 63.11% increase in the gender gap compared to the Pure Control group (p-value = 0.008). In terms of the equivalent number of GPA points, the resulting gender gap would equal a 2.1 GPA points increase (as opposed to 1.3 GPA in the pure control).

Since our goal is to understand whether demand for trust drives demand for women, the most meaningful subsample are employers that actually stated a preference for hiring women in their firms. In this subsample, the gender gap widens significantly more (+

113.24%), going from 6.76 percentage points in the Pure control, to 14.46 percentage points under Monitoring Behavior.[29]

Overall, the effect of Behavior audits on the gender gap are strongest for employers who state a preference for a nearly gender-balanced workforce composition. For this group, the gender gap is not statistically different from zero in the Control arm and rises to 15.22 percentage points (p-value < 0.001): this treatment effect is 2.7 times larger than the baseline estimate for the full sample.

In sum, the results are aligned with the predictions of our hidden discrimination hypothesis, in that the provision of monitoring support targeted at behavior and theft prevention widens the gender gap against women, revealing latent gender gaps especially among employers who appeared gender-neutral based on their stated preferences for hiring women.

A potential challenge to this interpretation is that monitoring audits may have indirect effects on hiring, particularly by deterring employers from hiring women on probation. For example, employers in the Monitoring-Behavior (MB) arm may reduce female hiring not because monitoring lowers the value of trust, but because they fear increased scrutiny could limit their opportunities for harassing women. The Monitoring-Safety (MS) arms, which we analyze next, provides a valuable benchmark because it allows us to test for scrutiny effects.

4.2.2 How do Safety audits affect the gender gap?

We compare the gender gap in hiring between employers in the Monitoring-Safety arm and those in the business-as-usual group (PC). As discussed in the design section, safety

[29] We exclude employers who report an ideal 100% male worker gender composition for whom there is little room to reduce the gender gap (ca. 15% of the sample). Because the gender gap is estimated at the level of individual randomization (CV-employer pair), restricting the sample to employers who express a preference for a mixed-gender workforce does not pose an econometric concern.

audits may or may not differentially affect female hiring. The direction of the effect, if there, is theoretically ambiguous, as two opposing mechanisms could be at play: a scrutiny effect that reduces the relative demand for women or a reassurance effect — from improved perceptions of workplace safety and reduced harassment concerns — that may increase the relative demand for women.

We find that Safety audits increase the relative demand for female trainees, with little effect on male trainees' selection rates, as shown in the raw data (Figure 5). This implies a reduction in the gender gap against women. Indeed, regression estimates of β_5 in Table 4, column (1), show that assignment to the Monitoring-Safety treatment reduces the gender gap by 8.6 percentage points on average. We can reject the null hypothesis that safety audits widen the gender gap at the 1% level (p-value = 0.005).

The patterns are consistent with safety support influencing hiring primarily through a reassurance channel, rather than by increasing scrutiny of employer behavior. Because these results isolate the effect of audits that target safety directly, they offer a useful benchmark for assessing whether indirect scrutiny effects might confound the impact of Monitoring Behavior audits. Given the evidence, we find it unlikely that such indirect effects are driving the Monitoring Behavior results.

We also note that even relatively mild safety support leads to a substantial reduction in the gender gap, an 83.5% reduction relative to the business-as-usual condition (p-value = 0.008). In fact, we cannot reject the null that the gender gap in the safety support arm is statistically indistinguishable from zero.

Thus, addressing frictions related to safety concerns meaningfully reduces observable gender gaps in employers' demand for workers. In contrast to the Behavior audits, the effects of safety audits are not concentrated among employers with particular preferences for a gender-mixed workforce. Instead, the impact is consistent across the preference distribu-

tion, with one exception: for employers stating no preference for hiring women, the audit has no impact on the gender gap (see Table 4, column 3).

5 Mechanism Discussion and Supporting Evidence

Our results align with the predictions of the hidden discrimination hypothesis. Reducing the costs of monitoring for workers' theft and behavior widens the gender gap. These results, together with the treatment effect of Safety audits that rule out scrutiny effects, support an interpretation such that easing monitoring constraints reduces the value employers place on trustworthiness, a trait stereotypically associated with female workers.

An important finding is that reducing employers' monitoring-behavior costs significantly widens the gender gap, especially among those who initially expressed the strongest demand for female workers. This suggests that the relatively mild gender gaps observed under business-as-usual, particularly among employers with the strongest stated preferences for hiring women, conflate a preference for diversity with a demand for trustworthiness. A simple back-of-the-envelope calculation indicates that the demand for trustworthiness explains at least 57% of the correlation between stated diversity preferences and the gender gap under business-as-usual conditions. Within the framework of standard discrimination models, the behavior of most employers—particularly those who state a preference for hiring women—is consistent with positive statistical discrimination in favor of women based on perceived trustworthiness. We draw on lab-in-the-field behavioral evidence from our main sample of employers and trainees to support our interpretation, and to assess the extent to which belief-based discrimination reflects inaccurate perceptions (Bohren et al., 2023).

5.0.1 Gender and Trustworthiness

We measured pro-sociality and trustworthiness of trainees in an in-person survey implemented at the VTIs. We complemented the trainee results with questions eliciting employers' beliefs about trainees' responses.

We elicited three complementary measures: i) survey answers on patience, altruism, positive reciprocity and trust using the experimentally validated Global Preference Survey (GPS), developed by Falk et al. (2018); ii) standard dictator, trust and cooperation games; iii) a novel lab-style task to measure trustworthiness and cheating behavior that builds on Caria and Falco (2024).[30]

Across these measures, we find no evidence of systematic gender differences in either skills or trustworthiness, and only mild evidence that male trainees exert more effort in tasks (Table A5, panel (a)). Looking at prosocial behavior, we find only minor gender differences in both the GPS answers and behavior in the standard games. Female trainees score significantly higher on positive reciprocity in the GPS measures (0.37 standard deviations; p-value = 0.010), consistent with employers' beliefs. They also score higher in altruism (0.29 standard deviations; p-value = 0.094), though employers expect the opposite. We find no significant gender differences in patience or generalized trust, despite employers strongly believing that women are more trusting (0.48 standard deviations, p-value < 0.001). Results from standard games partly diverge from GPS findings: in the Dictator Game, there are no significant gender differences in sharing, while employers

[30] We randomly assigned trainees to either a supervised or unsupervised condition. In the unsupervised condition, trainees received a flat wage to complete up to 12 proofreading tasks, with no verification of actual output. In the supervised condition, trainees were paid a piece rate, with supervisors verifying output and paying only for completed tasks. Employers were then asked to predict the average effort of male and female trainees under both conditions, allowing us to compare actual effort to employer beliefs. We introduced a measure of cheating: trainees could increase their pay by 50% by overstating their output. The sample was stratified by treatment and gender to estimate mean outcomes separately for each group. To avoid deception, individual output was not verified in the unsupervised arm. Misreporting could only be assessed at the treatment-by-gender cell level.

expect women to share less. In the Trust Game, women are significantly more likely to distrust, but just as likely as men to reward trust.

Results of the Supervision and Honesty game are described in columns (7)-(11) of Table A5 and Figures A10 and A9. Male trainees completed more tasks than female trainees in both supervised (p-value = 0.005) and unsupervised (p-value = 0.039) settings, while supervision itself had no significant effect on output (p-value = 0.516). Male trainees misreport less than female trainees in both the supervised (p-value = 0.068) and unsupervised (p-value = 0.066) conditions. Overall, employers tend to underestimate trainees' effort and overestimate the impact of supervision on increasing it. In particular, employers believe that male trainees will exert less effort than female trainees when unsupervised, and that this gap will disappear under supervision. While supervision significantly reduces misreporting overall (by 63%, p-value = 0.009); however, there is no differential effect by gender (p-value = 0.415). Employers expect male trainees to misreport task completion 11.5 percentage points more often than female trainees; under supervision, they expect this gap to shrink to 5.7 percentage points — but they do not anticipate a change in the relative gap.

6 Hidden Gender Bias?

Are employers accurately assessing trainees' trustworthiness and ability by gender? To what extent does the hidden gender gap reflect bias rather than pure safety concerns? To answer these questions, we complement employer experimental assessments with a survey of 182 vocational trainees measuring objective skills, prosocial behavior, and trustworthiness through various metrics including incentivized behavioral games (Appendix Table A14).

We begin by examining how gender influences employers' experimental perceptions of key trainee qualities, including work quality, trustworthiness, expected effort, and earnings potential — all else equal. Consistent with survey evidence, female trainees are perceived to be more trustworthy and honest (0.06 standard deviations, p-value = 0.040) but rated significantly lower than equivalent male trainees on perceived work quality (–0.17 standard deviations) and expected effort (–0.21 standard deviations) (see Appendix Table A6).[31]

According to this evidence, the gender gap in probationary hiring may be explained by statistical discrimination, where employers use gender as a proxy for ability. Following standard approaches in discrimination research (Bertrand and Duflo, 2017), we exploit randomized high-ability signals (e.g., high GPA) embedded in résumés to test employers' responsiveness to ability by gender. Specifically, we estimate:

$$\text{Meet}_{ijs} = \beta_0 + \beta_1 \text{Female}_{ij} + \beta_2 \text{HighAbility}_{ij} + \beta_3 \text{Female}_{ij} \cdot \text{HighAbility}_{ij} + \delta_i + \sigma_s + u_{ijs} \tag{3}$$

where HighAbility_{ij} indicates whether résumé i presented to employer j has high-ability signals. The coefficient of interest is β_3. We cluster standard errors at the employer and profile level.

Our primary ability measure is the GPA score, visually represented on résumés by 5 gold stars for top scorers and 3 stars for average scorers within vocational training cohorts. Additionally, we construct a résumé quality index combining indicators like secondary education completion, possession of a Directorate of Industrial Training (DIT) certification, enrollment in a 12-month program, and high GPA, all weighted equally.

[31] These variables were pre-registered as secondary outcomes.

We find no evidence that differences in worker ability or quality explain the gender gap, suggesting that statistical discrimination is not the main driver of disparities in probationary hiring. Column (1) of panel (b) in Table 3 shows that top-ranked CVs are 11.8 percentage points (25.4%, p-value < 0.001) more likely to be selected, but the interaction with gender is negative and statistically insignificant (p-value $= 0.307$), indicating female candidates do not benefit more from high-ability signals. The results are similar using the résumé ability index as shown in Table A3: a one-standard deviation increase in résumé quality increases the likelihood of selection by 13.2 percentage points (p-value < 0.001) but does not reduce the gender gap (p-value $= 0.454$).

We also find no evidence of gender differences in trainee ability or skills in our survey data. To measure skills, we ask trainees to self-report their ability to perform key technical tasks (such as "conducting an oil change" in mechanics) and to complete a set of theoretical questions specific to their specialization.[32] In self-assessments, presented in column (6), panel (a) of Table A14, female and male trainees are equally likely to report being able to perform a standard task in their sector (p-value $= 0.831$). On theory-based questions, women score 10.5 percentage points higher on average, though this difference is also not significant (p-value $= 0.249$).[33] However, employers systematically believe that female workers are less skilled, expecting them to perform worse on both practical and theoretical tasks.

Does ability explain the increase in the gender gap under the Monitoring Behavior intervention? We find no evidence that differences in worker ability or quality account for the widening gender gap. Table A4, panel (a), and Table A5, panel (a), report estimates for

[32]The ability tasks are adapted from Alfonsi et al. (2020), implemented in the same context. The theoretical questions were designed by VTI instructors; an example is shown in Figure A4. To match trainees' outcomes with employers' beliefs, we randomly selected one of seven theory exam questions and asked employers to evaluate that specific question.

[33]Regression results are summarized in the Appendix Table A14. The regressions include VTI and sector fixed effects, with standard errors clustered at the VTI level.

the GPA score measure and the standardized ability index, respectively. These results indicate that statistical discrimination is not the main driver of the disparities in probationary hiring induced by MB audits. Panel (b) shows similar patterns hold under the Monitoring Safety arm.

Although employers hold (inaccurate) gender-based priors on ability, we find no evidence that they use them in a way consistent with statistical discrimination. Under all monitoring intereventions, employers respond to ability overall, but not in a gender-differentiated way, suggesting that statistical discrimination does not explain the hidden gender gap under reduced monitoring costs.

Finally, while part of the hidden gap can be attributed to employers' safety concerns about female workers, a substantial share — approximately 40% — remains unexplained and is consistent with gender bias.

7 Implications for Policy

Consistent with Gentile et al. (2023), our paper identifies significant demand-side obstacles—including employer biases and preferences—that constrain female employment and can undermine the effectiveness of training interventions. This highlights the necessity of policies that address not only supply-side factors, but also employer attitudes and recruitment practices to achieve gender-equitable labor market outcomes.

Our key contribution, echoing Bandiera et al. (2022), shows that market conditions shape the success of interventions by influencing employer attitudes toward women. These interactions complicate predicting the effectiveness of training interventions across different settings. Therefore, it is important to think of the discrimination implications of broader

firm or labor market interventions. For instance, while reducing information asymmetries at the firm can lower monitoring costs, it may simultaneously erode women's perceived trust advantage, reinforcing gender segregation. This underscores the importance of complementary policies aimed at tackling underlying gender biases alongside market frictions.

While the trust premium facilitates women's entry into male-dominated sectors, structural barriers—such as search frictions and safety concerns—must be addressed to sustainably expand opportunities, consistent with theories of paternalistic discrimination (Buchmann et al., 2023). On the positive side, our evidence shows that relatively simple safety interventions can yield large gains not only by increasing women's labor supply, resonating with recent findings on transport safety (Field and Vyborny, 2022), but also by stimulating labor demand. Safety audits, frequently promoted by institutions such as the ILO (ILO, 2019), have emerged as a practical and scalable policy tool to improve workplace conditions for women.

Finally, shifting hiring practices from informal, referral-based systems common in smaller firms to more structured recruitment processes prevalent in larger firms can broaden candidate pools and enhance women's labor market participation (Alfonsi et al., 2020; Beaman et al., 2018).

8 Conclusions

We study gendered hiring in three male-dominated sectors in Uganda and show that market constraints shape both the extent and visibility of discrimination. Under the status quo, employers report a preference for gender-mixed workforces, leading to demand for female workers — when they are available — and a relatively small observed gender gap.

We hypothesize that this apparent demand for diversity masks a demand for trust, as women are perceived to have a comparative advantage in trustworthiness. Consistent with this, when employers' monitoring constraints are relaxed, the gender gap in probationary hiring widens significantly — especially among those who report stronger preferences for hiring more women at baseline. This gap reflects hidden gender discrimination, as we find no actual differences in ability or trust between male and female trainees.

The results align with Becker's taste-based discrimination framework: employers face a "discrimination cost" when acting on their bias against hiring women, and this cost increases with the perceived value of female candidates relative to male candidates.

References

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

_ , Oriana Bandiera, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali, "Tackling youth unemployment: Evidence from a labor market experiment in Uganda," *Econometrica*, 2020, *88* (6), 2369–2414.

Ashraf, Nava, Oriana Bandiera, Virginia Minni, and Víctor Quintas-Martínez, "Gender gaps across the spectrum of development: Local talent and firm productivity," 2024.

Bandiera, Oriana, Robin Burgess, Anton Heil, and Gabriel Leite Mariante, "Job diversification and economic growth," Technical Report, Technical Report, London School of Economics and Political Science 2022.

Bassi, Vittorio, Jung Hyuk Lee, Alessandra Peter, Tommaso Porzio, Ritwika Sen, and Esau Tugume, "Self-employment within the firm," Technical Report, National Bureau of Economic Research 2023.

Beaman, Lori, Niall Keleher, and Jeremy Magruder, "Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi," *Journal of labor economics*, 2018, *36* (1), 121–157.

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, 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.

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

Breza, Emily and Supreet Kaur, "Labor Markets in Developing Countries," *Annual Review of Economics*, 2025, *17*.

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.

Caria, Stefano A and Paolo Falco, "Skeptical employers: Experimental evidence on biased beliefs constraining firm growth," *Review of Economics and Statistics*, 2024, *106* (5), 1352–1368.

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

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

Eberhardt, Markus, Giovanni Facchini, and Valeria Rueda, "Gender differences in reference letters: Evidence from the economics job market," *The Economic Journal*, 2023, *133* (655), 2676–2708.

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.

Field, Erica and Kate Vyborny, "Women's mobility and labor supply: experimental evidence from Pakistan," *Asian Development Bank Economics Working Paper Series*, 2022, (655).

Gassier, Marine, Lea Rouanet, and Lacina Traore, "Addressing Gender-Based Segregation through Information," 2022.

Gentile, Elisabetta, Nikita Kohli, Nivedhitha Subramanian, Zunia Saif Tirmazee, and Kate Vyborny, "Barriers to entry: Decomposing the gender gap in job search in urban Pakistan," Technical Report, ADB Economics Working Paper Series 2023.

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

Goldin, Claudia D, "The role of World War II in the rise of women's employment," *The American Economic Review*, 1991, pp. 741–756.

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).

Hammond, Sarah, "The Secret History of the Female Code Breakers Who Helped Defeat the Nazis," *Politico Magazine*, October 10 2017.

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.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

ILO, "Violence and Harassment Convention, 2019 (No. 190)," https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO:12100:P12100_INSTRUMENT_ID:3999810 2019. Adopted at the 108th International Labour Conference, Geneva, 21 June 2019.

International Labour Organization, "ILO Modelled Estimates and Projections database (ILOEST)," https://ilostat.ilo.org/data/ 2025. Accessed January 7, 2025.

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, 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.

Rose, Evan K, "The rise and fall of female labor force participation during World War II in the United States," *The Journal of Economic History*, 2018, *78* (3), 673–711.

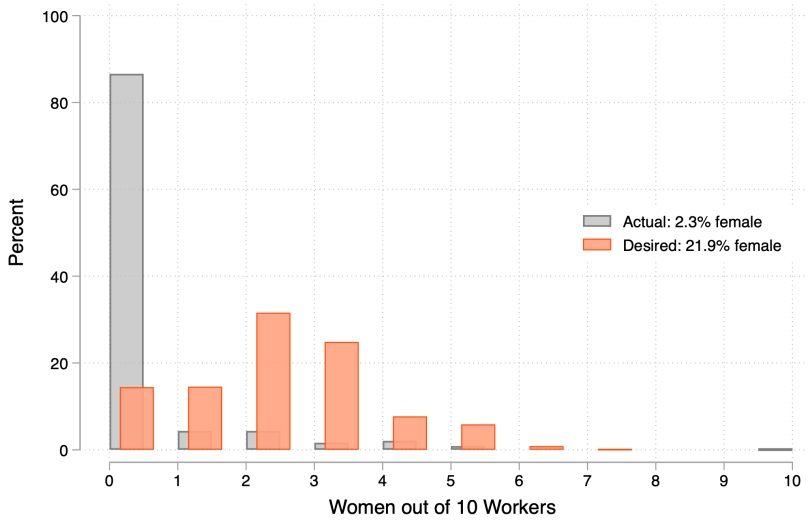
UBOS, "Census of Business Establishments, 2010/11," Technical Report, Uganda Bureau of Statistics 2011.

Uganda, DIT, "Internal Report," *Uganda Directorate of Industrial Training Internal Report*, 2023.

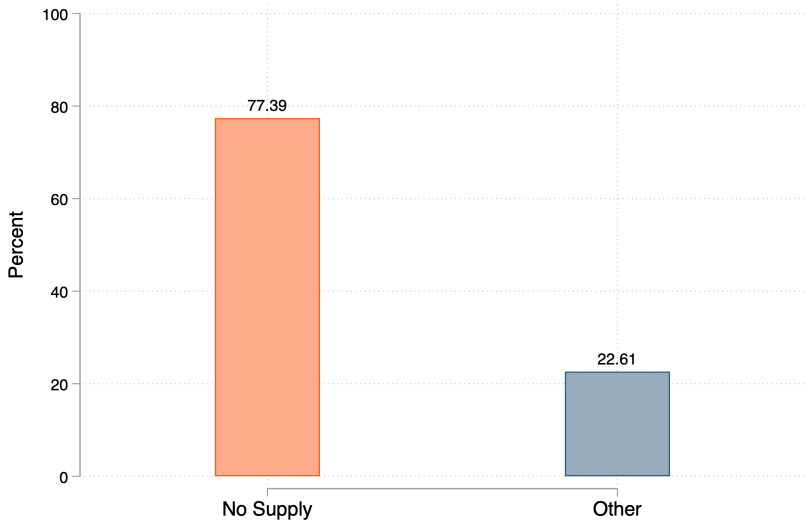
Figures

Figure 1: Employers' Gender Mix Preferences and Barriers to Hiring Women

(a) Employers' Desired and Actual Gender Mix Among Workforce



(b) "Can you explain why, in your opinion, you hire no female workers?"



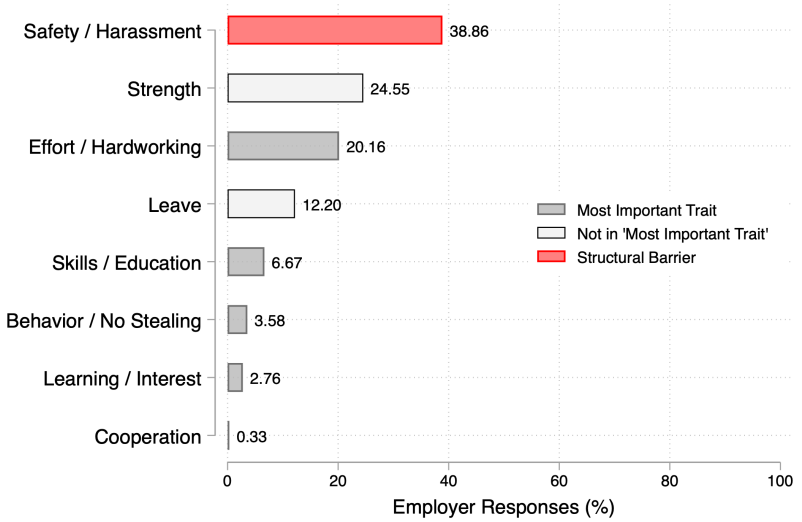
Notes: Panel (a) is based on all surveyed employers ($N = 921$) and it presents the distributions of their desired gender composition — on a scale from 0 men out of 10 (all women) to 10 men out of 10 (all men) — and the actual share of women among current workers. Panel (b) shows reported supply-side frictions in female hiring, based on data from employers who currently do not employ any female workers ($N = 795$). The data is derived from standardized responses to the question: " Can you explain why, in your opinion, you hire no female workers?". "Other" includes responses related to the following categories, listed by decreasing percentage: "Leave", "Strength", "Effort / Hardworking", "Skills / Education", "Safety / Harassment", "Behavior / No Stealing", and other residual reasons, often reflecting paternalistic views.

Figure 2: Gendered Perceptions of Valued Workers' Skills and Supply-Side Frictions

(a) Hardest-to-Find/Monitor Traits and Female Workers Better At

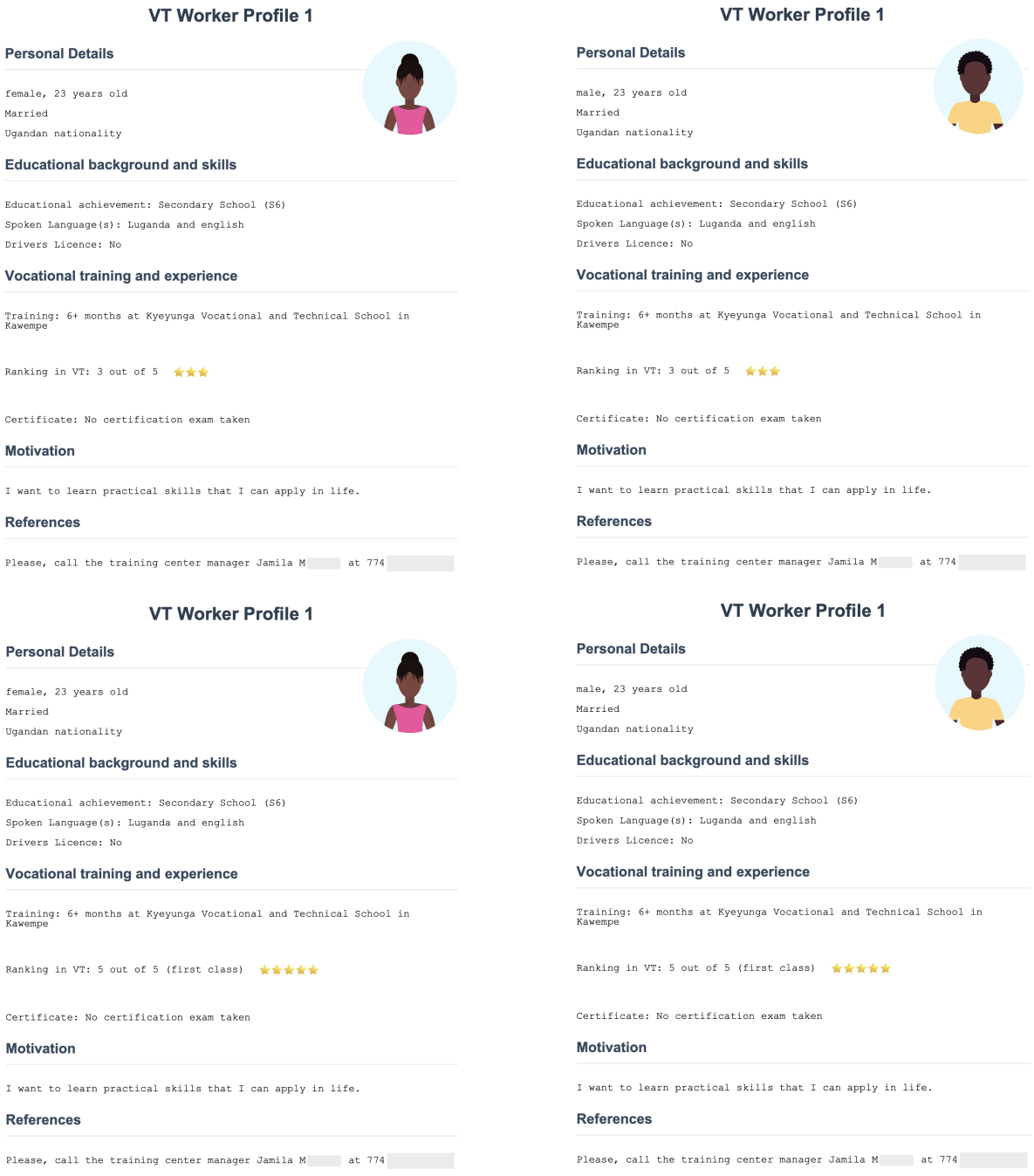


(b) "If you think of hiring a woman, what do you worry about?"



Notes: Panel (a) is based on all surveyed employers ($N = 921$) and reports in grey their responses to: "Of the following worker traits, which are the hardest to find and monitor?" in order from the most commonly mentioned (Behavior/No Stealing) to the least commonly mentioned (Cooperation). In purple, we present the share of respondents who answer "Female" to: "Do you think male or female workers are better at [trait]? Female, Male, or Indifferent.". The trait options come from responses re-coded into standardized categories, based on the open-ended question: "What is the single most important trait in a worker?". Panel (b) is based on employers who currently do not employ any female workers ($N = 795$) and it shows the distribution of their responses categorized from answers to the open-ended question:"If you think of hiring a woman, what do you worry about?". Of these, 10.73% responded "Nothing" and 1.79% "Other".

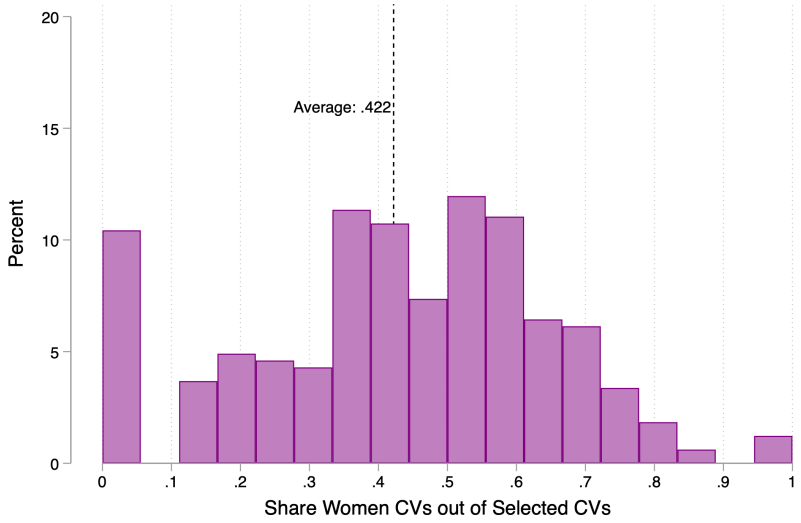
Figure 3: CV Examples



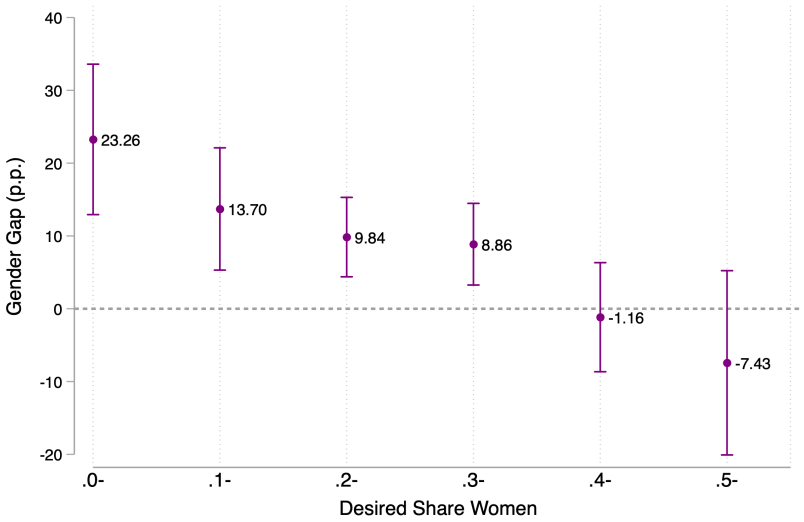
Notes: The figure presents examples of résumés in both female and male versions and for high and low ability levels. The content of the profile is randomized from 472 real applicants' data. Each résumé variation in ability is indicated by a GPA score out of 5 and represented with five stars. In total, there are 24 résumés, with four variations for each profile that reflect the 2×2 design based on gender and ability.

Figure 4: Proportion of Female résumés Selected by Employers

(a) Distribution



(b) Gender Gap by Stated Preferences for Workforce Gender-Mix

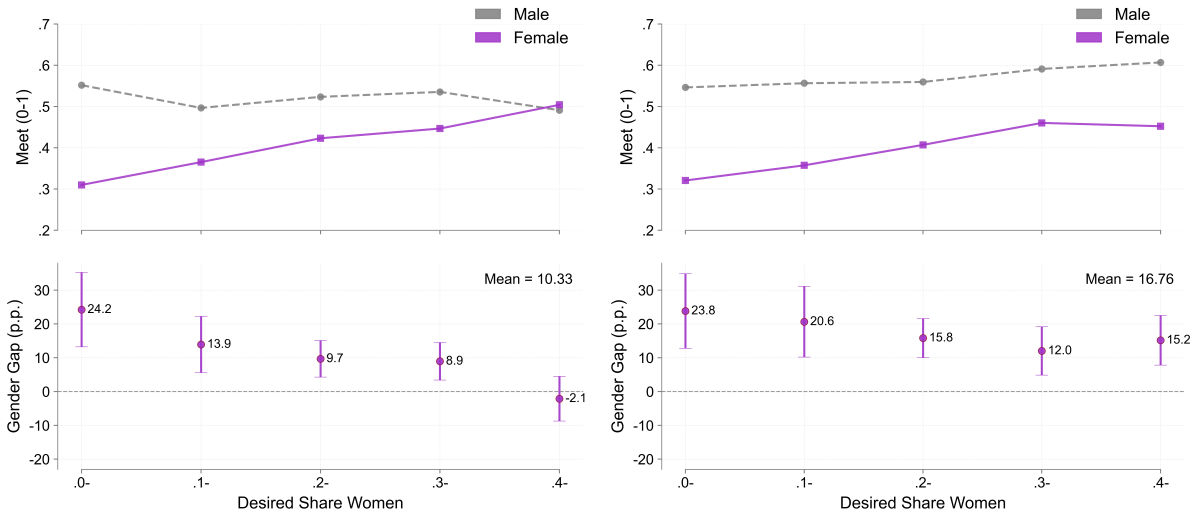


Notes: The experimental data comprises 7,365 hiring decisions made by the 326 employers in the Pure Control group (PC). Panel (a) illustrates the distribution of the share of female candidates selected over the total number of candidates selected for a meeting, with an average of 42.2% of female workers. Panel (b) presents the gender gap estimates for the request to meet candidates by stated preferences for workforce gender-mix. Estimates are obtained by regressing a meeting request indicator on a female candidate dummy on the sub-sample of employers at a given level of preferences for workforce gender-mix. Employers with preferences for an equal-gender distribution or more female than male workers are grouped. Regressions include strata fixed effects (since randomization was stratified by sector, enumerator, survey date, and profile). The bars represent the 95% confidence interval. Standard errors are clustered at the employer and profile level.

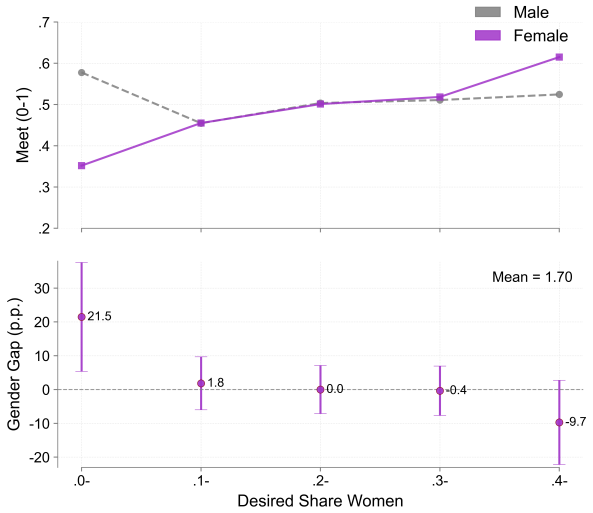
Figure 5: Gender Gap in Hiring by Stated Preferences for Workforce Gender Mix and Monitoring Treatment

(a) Pure Control (PC)

(b) Monitoring Behavior (MB)

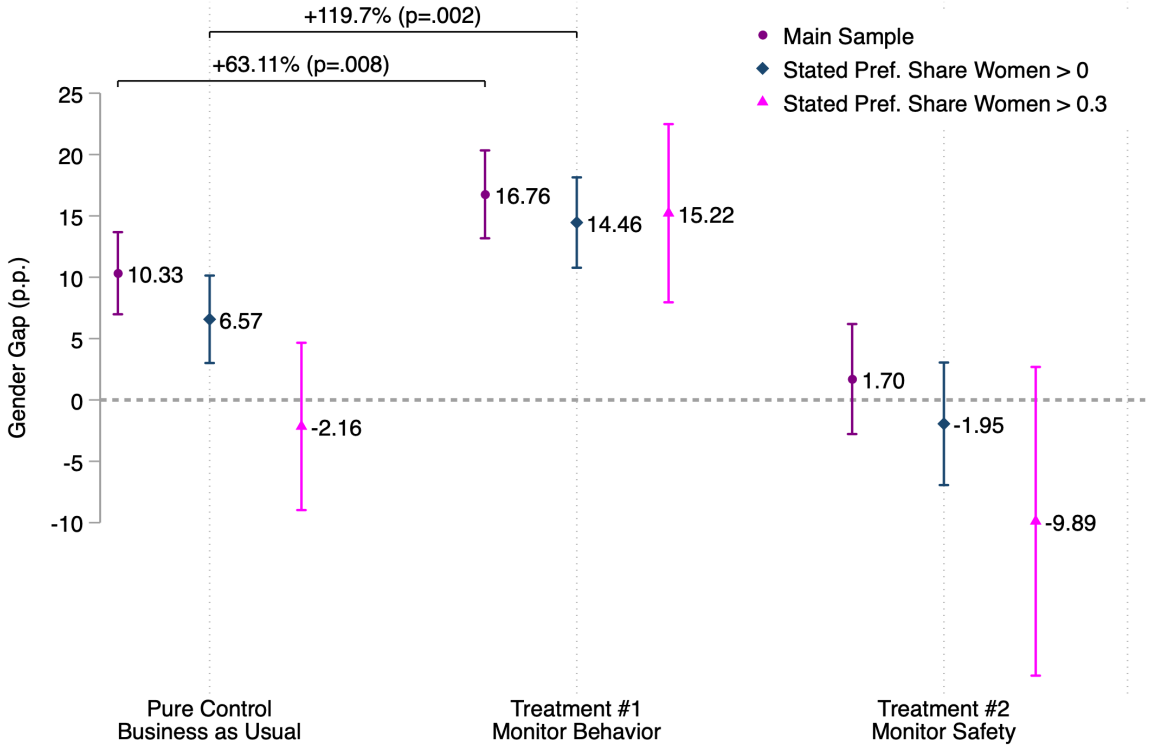


(c) Monitoring Safety (MS)



Notes: The figures show the heterogeneity in the gender gap by stated preferences for workforce gender mix, by treatment arm. Panel (a) presents descriptive statistics for the Pure Control group, while panels (b) and (c) present the corresponding statistics for the Monitoring Behavior and Monitoring Safety groups, respectively. The top graph of each panel presents the raw data average of the employer's interest in meeting the candidate by profile gender by the share of female workers preferred at baseline, while the bottom graph displays the heterogeneous treatment effect of a résumé being assigned to a female candidate, estimated with our main specification from Equation 2. Confidence intervals are at the 95% level.

Figure 6: Gender Gaps by Monitoring Arm and Preferences for Workforce Gender-Mix



Notes: The figure presents results from the main experiment. Point estimates and 95% confidence intervals for the gender gap against women are estimated using the regression model in Equation 2, and are reported by treatment arm. In the Pure Control group, employers evaluate candidate profiles under business-as-usual conditions. In the Monitor Behavior group, employers receive audit visits to monitor the behavior of new workers. In the Monitor Safety group, employers receive monitoring support focused on worker safety. The figure presents experimental results across three levels of employers' stated preferences for workforce gender mix. The purple circles represent estimates from the main specification (Equation 2) ($N = 20,726$). The blue squares represent estimates from employers with any stated preferences for workforce gender mix ($N = 17,748$). The regression uses the main sample and includes an indicator variable for whether the employer has any stated preferences for workforce gender mix, fully interacted with both the gender of the candidate and the randomly assigned monitoring condition. The orange triangle represents estimates from the same regression model but using an indicator for employers in the top quintile of the stated preferences for workforce gender mix (i.e., those preferring more than 30% female workers; $N = 3,031$). All regressions include strata fixed effects. Standard errors are clustered at both the individual and profile levels.

Tables

Table 1: Descriptive Statistics by Sector

	All	Mechanics		Carpentry		Welding	
	Mean	Mean	[Q25;Q75]	Mean	[Q25;Q75]	Mean	[Q25;Q75]
Panel A: **Firms**							
N. employees	8.34	15.16	[7 ; 20]	4.55	[2.5 ; 5]	5	[3 ; 6]
Any female employee	0.14	0.13	[0 ; 0]	0.19	[0 ; 0]	0.09	[0 ; 0]
N. female employees	0.19	0.19	[0 ; 0]	0.26	[0 ; 0]	0.12	[0 ; 0]
N. trainees	2.64	5.4	[2 ; 6]	0.97	[0 ; 2]	1.41	[0 ; 2]
Share family empl.	0.14	0.07	[0 ; 0.12]	0.16	[0 ; 0.25]	0.2	[0 ; 0.33]
Age	10.73	12.11	[6 ; 16]	10.01	[5 ; 14]	9.99	[5 ; 13]
Firm capital (USD)	1215.0	1594.4	[0.04 ; 1852]	933.4	[0.04 ; 1037]	1079.1	[0.04 ; 1481]
Firm revenues (USD)	718.1	589.3	[0.04 ; 926]	705.0	[0.04 ; 1111]	874.1	[0.04 ; 1407]
Firm profits (USD)	286.0	246.2	[0.04 ; 370]	276.8	[0.04 ; 370]	339.4	[0.04 ; 556]
Daily n. customers	11.76	6.44	[4 ; 8]	14.37	[6 ; 20]	14.76	[5 ; 20]
Any stealing	0.83	0.9	[1 ; 1]	0.81	[1 ; 1]	0.77	[1 ; 1]
Monitor constrained	0.9	0.89	[1 ; 1]	0.93	[1 ; 1]	0.89	[1 ; 1]
Panel B: **Employers**							
Male	0.96	0.96	[1 ; 1]	0.96	[1 ; 1]	0.97	[1 ; 1]
Age	37.08	40.08	[32 ; 47]	35.78	[28 ; 42]	35.24	[29 ; 40]
Years of experience	15.21	18.66	[10 ; 25]	13.19	[6 ; 19]	13.6	[8 ; 20]
Higher education	0.63	0.64	[0 ; 1]	0.62	[0 ; 1]	0.65	[0 ; 1]
VTI	0.31	0.37	[0 ; 1]	0.27	[0 ; 1]	0.29	[0 ; 1]
Daily hours worked	10.44	10.47	[10 ; 12]	10.48	[9 ; 12]	10.38	[10 ; 11]
Daily hours monitor	2.06	2.03	[1 ; 2]	2.1	[1 ; 2]	2.07	[1 ; 2]
Women accepted	0.66	0.59	[0 ; 1]	0.61	[0 ; 1]	0.78	[1 ; 1]
Risks for women	0.53	0.25	[0 ; 0]	0.58	[0 ; 1]	0.76	[1 ; 1]
Daily extra monitor	0.65	.	[. ; .]	0.44	[0 ; 1]	0.76	[0 ; 1]
Observations	921	318		300		303	

Notes: The table reports baseline characteristics for all surveyed firms, with means and interquartile ranges by sector. Panel A shows firm characteristics; Panel B shows employer characteristics. Monetary values in USD. *Any female employee* equals 1 if the firm employs at least one woman. *Share of family empl.* is the share of workers who are family or close friends. *Age* refers to the firm or the employer (in years). *Any stealing*, *Monitoring constrained*, *Higher education*, and *VTI* are binary indicators for experienced workers stealing, more monitoring required, completed secondary education, and VTI, respectively. *Daily extra monitor* is measured in extra hours of monitoring needed. *Women accepted* and *Risks for women* equal 1 if the respondent perceives female acceptance or risks at the workplace.

Table 2: Balance Table

	PC (1)		MB (2)		MS (3)		p-value	p-value	N
	Mean	SD	Mean	SD	Mean	SD	(1)-(2)	(1)-(3)	
Panel A: **Firms**									
N. employees	8.24	(8.96)	8.67	(8.66)	7.95	(9.14)	0.085	0.502	903
Any female employee	0.13	(0.34)	0.15	(0.36)	0.13	(0.34)	0.389	0.879	906
N. female employees	0.19	(0.55)	0.21	(0.56)	0.17	(0.48)	0.464	0.831	906
N. trainees	2.52	(3.43)	2.57	(3.25)	2.81	(5.48)	0.381	0.097	906
Share of family empl.	0.15	(0.25)	0.15	(0.23)	0.13	(0.22)	0.527	0.115	898
Age	10.62	(7.82)	11.27	(8.11)	10.35	(7.23)	0.244	0.671	906
Firm capital (USD)	1245	(2776.4)	1437	(3547.9)	974.7	(1421.8)	0.220	0.136	866
Firm revenues (USD)	649.2	(861.9)	699.9	(885.8)	805.6	(1055.4)	0.556	0.122	861
Firm profits (USD)	259.8	(396.4)	279.8	(405.2)	317.1	(468.7)	0.634	0.187	865
Monthly n. customers	11.9	(17.65)	12.03	(13.2)	11.1	(11.87)	0.810	0.109	906
Any stealing	0.83	(0.38)	0.85	(0.36)	0.81	(0.39)	0.667	0.528	906
Monitoring constrained	0.9	(0.31)	0.92	(0.28)	0.9	(0.3)	0.395	0.930	906
Panel B: **Employers**									
Male	0.97	(0.17)	0.95	(0.23)	0.98	(0.14)	0.115	0.387	906
Age	37.12	(9.84)	36.47	(9.25)	37.26	(10.26)	0.472	0.620	903
Years of experience	15.12	(8.96)	15.38	(9.11)	14.81	(9.01)	0.580	0.719	904
Higher education	0.62	(0.49)	0.62	(0.49)	0.66	(0.47)	0.966	0.491	906
VTI	0.27	(0.45)	0.34	(0.47)	0.32	(0.47)	0.058	0.159	906
Daily hours worked	10.55	(1.54)	10.45	(1.7)	10.35	(1.82)	0.592	0.237	906
Daily hours of monitoring	2.08	(1.58)	2	(1.29)	2.08	(1.52)	0.459	0.970	902
Women accepted	0.67	(0.47)	0.65	(0.48)	0.66	(0.47)	0.457	0.495	906
Risks for women	0.47	(0.5)	0.58	(0.49)	0.54	(0.5)	0.003	0.125	906
Daily extra monitor	0.51	(0.85)	0.76	(1.63)	0.67	(1.27)	0.146	0.253	301

Notes: The table presents the baseline balance for the main sample of the experiment. Columns (1), (2), and (3) report means and standard deviations by treatment arm: Pure Control (PC), Monitoring Behavior (MB), Monitoring Safety workplace (MS). Standard deviations are reported in parentheses in separate columns. The difference columns (1)-(2) and (1)-(3) report the p-values generated by a regression of each outcome on a treatment dummy with robust standard errors and strata fixed effects. Significance levels are based on randomization inference p-values: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The last column reports the sample size of the regression used to run the difference-in-means tests. Some values for specific variables are missing because of enumerator mistakes or because the respondent did not know the answer or refused to provide an answer. The variables are the same as in Table 1.

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

Panel A: **Gender Only**

			Stated Preferences for Workforce Gender Mix				
Meet (0-1)	(1) Main Sample	(2) Gender-Mix Pref. > 0	(3) Bottom QT	(4) 2nd QT	(5) 3rd QT	(6) 4th QT	(7) Top QT
Female	-0.103***	-0.077***	-0.233***	-0.153***	-0.098***	-0.083***	0.026
	[0.016]	[0.015]	[0.050]	[0.038]	[0.026]	[0.027]	[0.032]
Observations	7,365	6,206	1,159	870	2,340	1,764	1,232
R-squared	0.123	0.123	0.200	0.190	0.149	0.129	0.137
Control mean	0.522	0.517	0.552	0.497	0.523	0.535	0.491

Panel B: **Gender $\times$ Ability**

			Stated Preferences for Workforce Gender Mix				
Meet (0-1)	(1) Main Sample	(2) Gender-Mix Pref. > 0	(3) Bottom QT	(4) 2nd QT	(5) 3rd QT	(6) 4th QT	(7) Top QT
Female	-0.087***	-0.058**	-0.233***	-0.148***	-0.048	-0.083**	0.034
	[0.022]	[0.021]	[0.052]	[0.051]	[0.031]	[0.037]	[0.039]
Female x High Ability	-0.033	-0.042	0.004	-0.008	-0.103***	-0.002	-0.016
	[0.032]	[0.029]	[0.062]	[0.061]	[0.033]	[0.053]	[0.052]
High Ability	0.118***	0.121***	0.105**	0.121**	0.141***	0.117***	0.093*
	[0.018]	[0.019]	[0.044]	[0.053]	[0.025]	[0.040]	[0.047]
Observations	7,365	6,206	1,159	870	2,340	1,764	1,232
R-squared	0.134	0.134	0.212	0.204	0.160	0.142	0.144
Control mean	0.464	0.459	0.491	0.445	0.448	0.481	0.455

Notes: The table reports the treatment effects from estimation regression models 1 (Panel A) and 3 (Panel B) for the sample of employers in the Pure-Control group (PC), and by quintiles distribution of preferred gender workforce composition of the employer in PC. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintiles distribution of preferred gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants to meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a female candidate. *High-Ability* is a dummy indicating whether the résumé has a top GPA score (5/5) within their vocational training program. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Experimental Results — OLS Gender by Monitoring Arm

			Stated Preferences for Workforce Gender Mix				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meet (0-1)	Main Sample	Gender-Mix Pref. > 0	Bottom QT	2nd QT	3rd QT	4th QT	Top QT
Female	-0.103***	-0.078***	-0.235***	-0.158***	-0.099***	-0.087***	0.029
	[0.016]	[0.016]	[0.052]	[0.039]	[0.026]	[0.028]	[0.034]
Monitor Behavior	0.052***	0.061***	-0.014	0.038	0.052*	0.060**	0.104***
	[0.014]	[0.015]	[0.040]	[0.037]	[0.027]	[0.026]	[0.034]
Monitor Safety	-0.013	-0.017	-0.000	-0.061*	-0.016	-0.020	-0.005
	[0.015]	[0.017]	[0.047]	[0.035]	[0.026]	[0.027]	[0.044]
Female x Monitor Behavior	-0.064***	-0.077***	0.003	-0.050	-0.061	-0.040	-0.178***
	[0.022]	[0.023]	[0.071]	[0.061]	[0.038]	[0.044]	[0.047]
Female x Monitor Safety	0.086***	0.091***	0.026	0.137**	0.104**	0.091*	0.070
	[0.028]	[0.028]	[0.089]	[0.054]	[0.044]	[0.045]	[0.073]
Observations	20,725	17,748	2,977	3,002	6,471	5,244	3,031
R-squared	0.127	0.126	0.177	0.175	0.147	0.119	0.124
Control Mean	0.522	0.517	0.552	0.497	0.523	0.535	0.491

Notes: The table reports the treatment effects for the main sample, and by quintiles distribution of stated preferred gender workforce composition of the employer. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintiles distribution of preferred gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a woman. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A Appendix

A.1 Experimenter Demand Effects

A potential concern is that employers may have tried to please experimenters by stating stronger preferences for gender mix in the workforce, or hiring more women than they would in reality. To mitigate this concern, choices in our experiment are incentive compatible: employers receive referrals of real workers based on their choices in the experiment. This setup makes it costly to misrepresent preferences, following the incentivized ratings paradigm (Kessler et al., 2019; Buchmann et al., 2023; Bohren et al., 2022).

We also conducted an additional test to assess this potential bias and found no evidence of experimenter demand effects. In the flavor of De Quidt et al. (2018), we asked employers: "Do you think we would prefer you hire men, women, or does it not matter?". The majority (66.7%) said it would not matter, and the second most common answer (23.5%) was that we preferred men.[34]

In line with the high female labor force participation in the country, we find broadly favorable attitudes toward women's employment among employers. About 84% of all employers (including 96.4% of men) agree that women should be free to work outside the home, and interestingly 82.8% support prioritizing women for leadership roles (we use standard World Values Survey questions). Yet, we could only identify one affirmative action policy relevant to employers—the UNDP's "Gender Equality Seal" for Ugandan enterprises by the United Nations Development Programme and awareness of it was very low: only 18.7% of employers had heard of it.

A.2 Within Subject Design

We extended the design in the welding and carpentry sectors (approximately two-thirds of the sample) by adding a *within-subject* component: after completing the initial 24 ratings under the main treatment condition, each employer evaluated 12 additional profiles under a second monitoring condition. We amended the pre-registration to reflect this change. This design is not our primary specification, but it offers greater statistical power. The results are consistent in sign and similar in magnitude to those obtained in the between-subjects design.

The results in the within-subject specification, involving 64.8% of the sample of surveyed employers, are consistent with our main specification in the across-subject design. Providing an employer with monitoring support reduces demand for female workers and increases the gender gap in hiring against women by 100.1% (12.2 percentage points increase, p-value 0.000) compared to a business-as-usual condition (Table A12, column (1)).

A.3 Differences From Preregistration

We note three main deviations from the original pre-analysis plan. First, the second primary outcome was not recorded as initially intended due to a coding mistake in the survey.[35] Second, we expanded the study to include two additional sectors — carpentry and welding — in addition to mechanics, which was

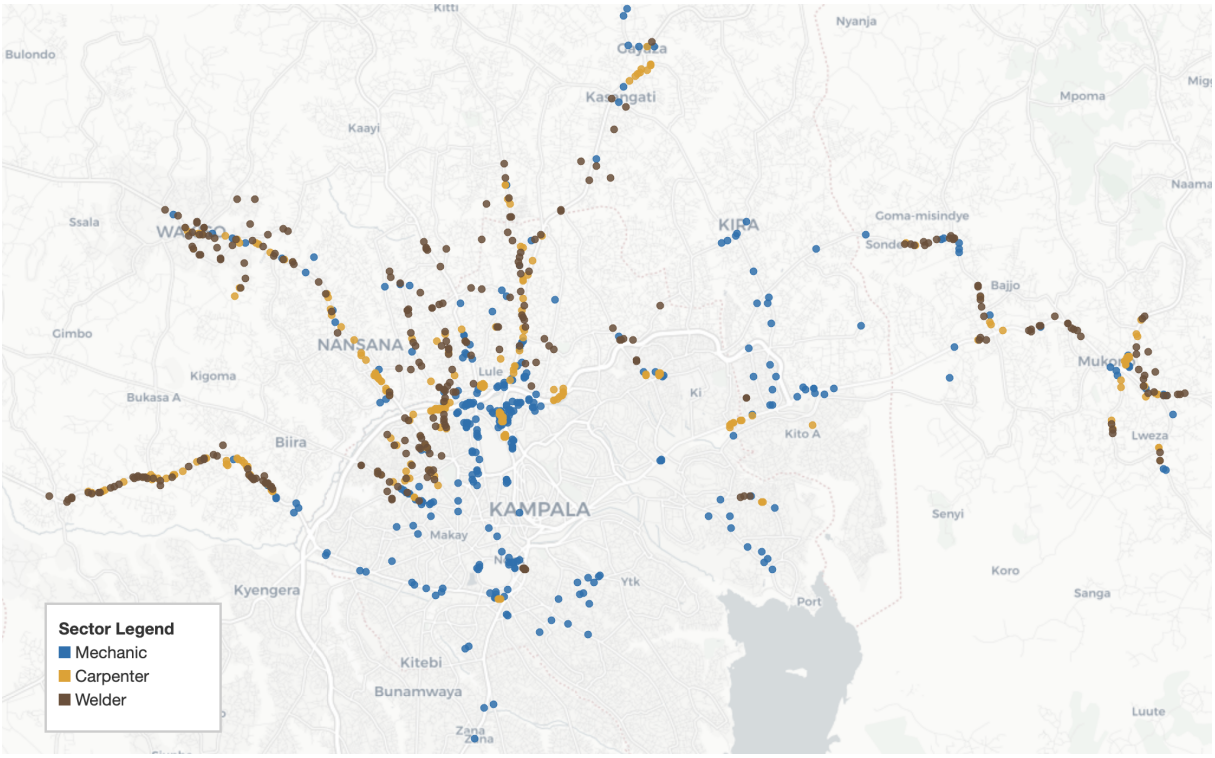
[34] Only 6.8% believed we preferred women. The remaining 3% refused to answer.

[35] *offer*: "How likely would you be to offer this worker a position as a mechanic at your firm? Please rate on a scale from 0 to 10, where 0 is very unlikely and 10 is very likely."

the only sector originally planned due to initial funding constraints. Third, while the safety audit arm was initially framed as an active control, we now consider it as an additional treatment of interest, given its distinct mechanisms and potential to influence hiring behavior.

A.4 Appendix Figures

Figure A1: Firms Spatial Distribution



Notes: This figure reports the locations of firms in the study sample in the metropolitan area of Kampala, Uganda. For privacy reasons, locations accuracy has been reduced by showing a random positions within 100m radius of the actual location.

Figure A2: Girls With Tools Trainees (Automotive)

(a) Practical classroom session.



(b) Internship.



Photo credit: © Mariajose Silva-Vargas.

Figure A3: Firm Layout — An Example (Automotive)

(a) Establishment Layout.



(b) Employer's Supervision.



Figure A4: Theory Technical Exam — An Example (Automotive)

FULL NAME:_____ ☐ MAN

ID:_____ ☐ WOMAN

MOTOR-MECHANICS

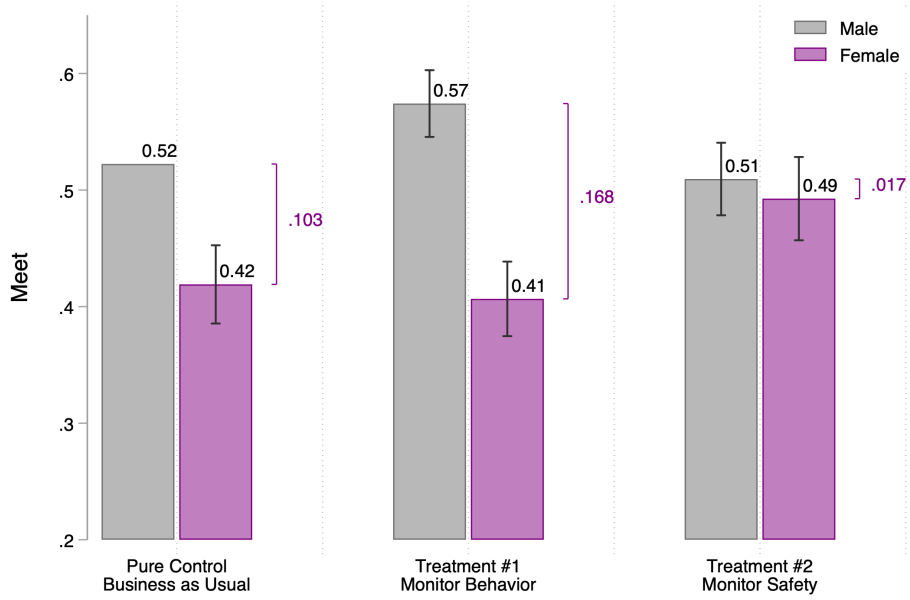
	Question	Answers
1	*multiple-choice* What are you advised to do when servicing the engine by changing oil?	A. Top up lubricating oil B. Replace oil filter C. Over hand engine D. Over hand cylinder head
2	*multiple-choice* What immediate remedy can you give to a vehicle with a problem of excessive tyre wear in the center more than other parts?	A. Increase tyre pressure B. Reduce tyre pressure C. Inflate pressure D. Remove the vehicle tire
3	*Multiple-choice* If a customer reports to you that his/her vehicle charging system works at lower rate, how can you help him?	A. Replacing the charging system B. Adjusting the alternator tension C. Replacing alternator housing D. Renewing wire insulator
4	*Multiple-choice* Which of the following set of systems or component call for mechanical adjustment during general vehicle service?	A. Tyres, cooling system, master cylinder B. Break shoes, alternator, and valve clearance C. Distributor, radiator, propeller shaft D. Tank, crank shaft, Turbo charger
5	*Multiple-choice* What solution would you give a customer with a vehicle engine producing blue smoke?	A. Top up lubricant B. Time the engine C. Replace piston rings D. Remove carbon deposits
6	*Matching* What should you do to stop the following vehicle troubles?	1. Battery over charging — A. Leaking fuel tank 2. Engine over heating — B. Renew regulator 3. Lubricant leakage — C. Reduce oil to the correct level 4. Smoke in exhaust — D. Renew piston rings 5. Engine fails to start — E. Charge the battery ____ ; ____ ; ____ ; ____ ; ____ ; ____ .
7	*Order* When changing engine oil, in which order should you perform the following steps?	A. Drain oil through drain plug B. Remove oil filter cup C. Run engine to check leaks D. Fill new oil through filler cup to level E. Remove oil filter F. Warm up the engine 1____ ; 2____ ; 3____ ; 4____ ; 5____ ; 6____ .

Figure A5: Valuation and Gender-Based Perceptions of Stated Most Important Worker's Skills



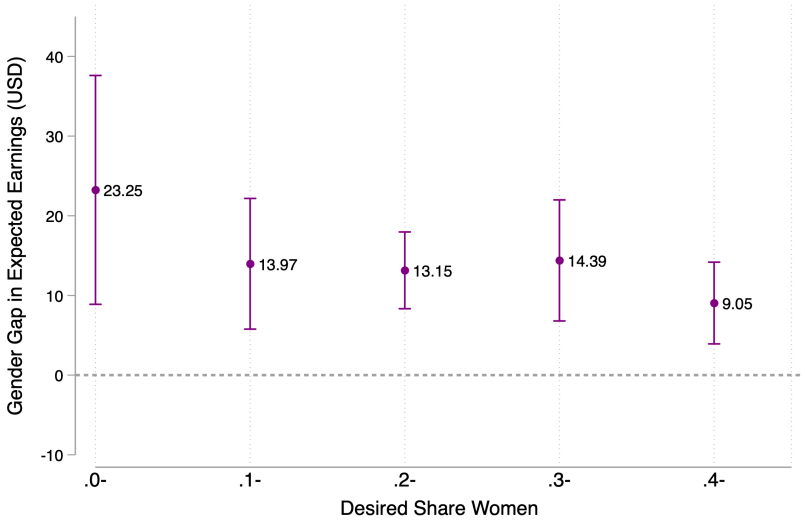
Notes: The figure presents descriptive statistics from the employers' baseline survey ($N = 921$). Panel (a) presents summary statistics on employers' reported importance of workers' traits, based on the open-ended question: "What is the single most important trait in a worker?". Responses were grouped into the standardized categories displayed in the figure. Of these, 13.43% responded "Other", a residual category encompassing answers often related to personal characteristics, such as workers' age or their personal background. Panel (b) presents the share of respondents who answer "Male" to: "Do you think male or female workers are better at [trait]? Female, Male, or Indifferent.". The trait options were derived from responses reported in Panel (a).

Figure A6: Experimental Results — OLS Gender by Monitoring Arm



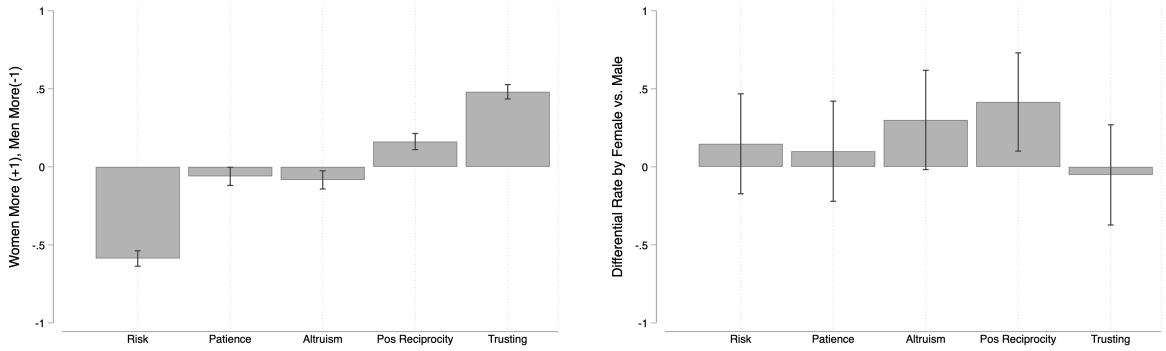
Notes: The figure reports the treatment effects for the main sample, and by quintiles distribution of preferred gender workforce composition of the employer. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a woman. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A7: Beliefs About Labor Market Outside Options



Note: Experimental data, résumé evaluations by employers in the control group ($N = 7,365$). The figure reports the predicted gender gap in monthly earnings (USD) of the candidates a year from the hire by stated preferences for workforce gender mix. The average gender gap is USD 14, with men expected to earn an average of USD 117.9 monthly (12.3%). Estimates are based on specifications that include fixed effects.

Figure A8: Gender Differences in Social Preferences



(a) Employers: Beliefs

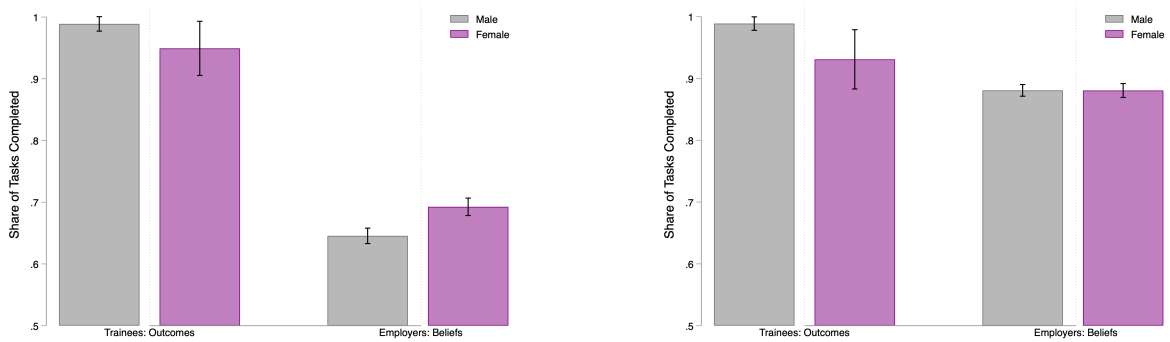
(b) Trainees: Outcomes

Note: Survey data survey to employers ($N = 920$) and to trainees ($N = 182$). We ask employers to predict the share of male and female trainees who 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."; and Trusting "I assume that people have only the best intentions.". All questions were asked on a scale from 0 (not willing/agreeing at all) to 10 (fully willing/agreeing). Panel (a) shows employers' beliefs; Panel (b) shows trainees' answers.

Figure A9: Gender Differences in Effort and Response to Supervision

(a) Unsupervised

(b) Supervised

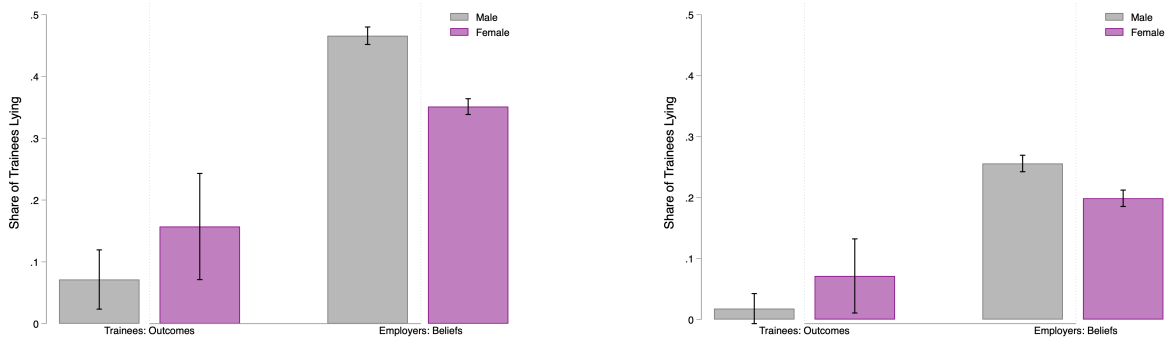


Notes: All surveyed employers were asked to predict the number of tasks that male and female trainees would complete under unsupervised (panel (a) and supervised (panel (b)) conditions, serving as a measure of effort ($N = 921$). In the unsupervised condition, the payment was made in advance as no verification could be made. In the supervised conditions, payment was piece-rate. We ask employers to estimate the number of tasks completed under both conditions, serving as a measure of effort. Each panel presents the share of tasks completed by trainees, disaggregated by gender, along with employers' beliefs about those shares.

Figure A10: Gender Differences in Lying and Response to Supervision

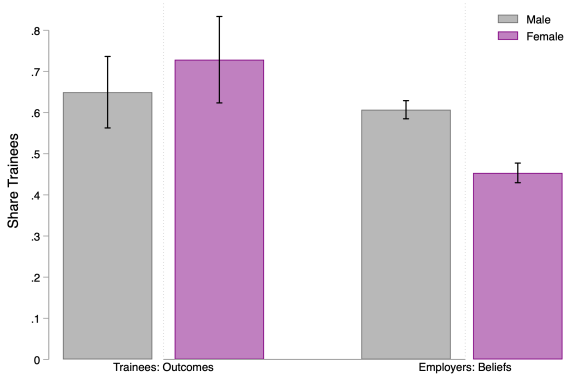
(a) Unsupervised

(b) Supervised



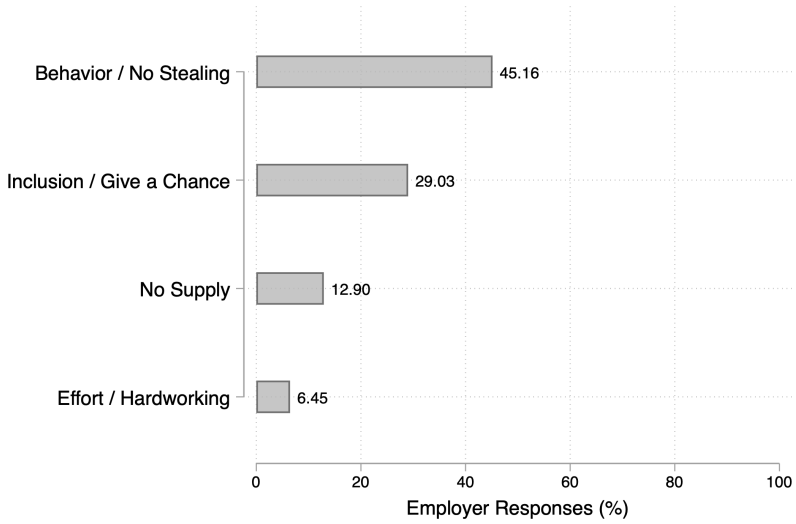
Notes: All surveyed employers were asked to predict the number of tasks that male and female trainees would misreport under unsupervised (panel (a) and supervised (panel (b)) conditions, serving as a measure of effort ($N = 921$). The task was designed to incentivize workers to misreport the number of tasks they had actually completed. Thus, we ask employers to estimate the extent of task misreporting under both conditions, serving as a measure of lying or misbehavior. Employers were asked to predict the number of male and female trainees who would misreport. Each panel presents the share of trainees lying about task completion, disaggregated by gender, along with employers' beliefs about those shares.

Figure A11: Gender Differences in Technical Skills



Notes: The data is from all surveyed employers ($N = 921$) and all surveyed trainees ($N = 182$). The figure reports the share of trainees stating to be able to perform a certain technical task (first two bars from the left) and the corresponding employers' beliefs about those shares (two bars on the right). The three questions are sector-specific. For mechanics: "Can you perform an oil change on a vehicle?"; For welders: ""; For carpenters: "What is the best practice for ensuring a smooth finish on a wooden surface?"

Figure A12: Experimental Female Hires in Male-Dominated Firms



Notes: The data is based on employers who expressed preferences for male workers only and currently do not employ any women, but selected at least one female CV in the experiment ($N = 32$). The figure presents their responses to the open-ended question: "Can you explain why you were choosing women in the exercise with the profiles before?". Responses were re-coded into standardized categories, with 9.68% falling under "Other". "No Supply" indicates question misunderstanding (e.g.: "There are no women workers that have come to ask for jobs").

A.5 Appendix Tables

Table A1: Descriptive Statistics

	Mean	SD	Min	Median	Max	N
Firm characteristics						
N. employees	8.34	8.9	1	5	60	918
Any female employee	0.14	0.34	0	0	1	921
N. female employees	0.19	0.53	0	0	3	921
N. trainees	2.64	4.12	0	2	52	921
Share family empl.	0.14	0.23	0	0	1	913
Age	10.73	7.75	0	9	47	921
Firm capital (USD)	1215.0	2727.3	0	555.6	35185.2	881
Firm revenues (USD)	718.1	938.5	0	370.4	5925.9	876
Firm profits (USD)	286.0	425.6	0	148.2	2963.0	880
Daily n. customers	11.76	15.19	1	7	180	921
Any stealing	0.83	0.38	0	1	1	921
Monitor constrained	0.9	0.29	0	1	1	921
Employer characteristics						
Male	0.96	0.19	0	1	1	921
Age	37.08	9.9	19	36	89	918
Years of experience	15.21	9.09	0.5	14	50	919
Higher education	0.63	0.48	0	1	1	921
VTI	0.31	0.46	0	0	1	921
Daily hours worked	10.44	1.68	6	10	24	921
Daily hours monitor	2.06	1.48	0	2	12	917
Women accepted	0.66	0.48	0	1	1	921
Risks for women	0.53	0.5	0	1	1	921
Daily extra monitor	0.65	1.29	0	0	11	302

Notes: The table reports baseline characteristics for all surveyed firms. Panel A shows firm characteristics; Panel B shows employer characteristics. Monetary values in Usd. *Any female employee* equals 1 if the firm employs at least one woman. *Share of family empl.* is the share of workers who are family or close friends. *Age* refers to the firm or the employer (in years). *Any stealing*, *Monitoring constrained*, *Higher education*, and *VTI* are binary indicators for experienced workers stealing, more monitoring required, completed secondary education and VTI, respectively. *Daily extra monitor* is measured in extra hours of monitoring needed. *Women accepted* and *Risks for women* equal 1 if the respondent perceives female acceptance or risks on the job. Some values for specific variables are missing because of enumerator mistakes or because the respondent did not know the answer or refused to provide an answer.

Table A2: OLS Attrition

	Monitor Treat		Female Treat
	(1) PC vs. MB Sample	(2) PC vs. MS Sample	(3) PC Sample
Main Sample	0.140	-0.092	-0.032
	[0.125]	[0.058]	[0.022]
Observations	14,954	14,916	7,812
R-squared	0.001	0.002	0.000
p-value	0.264	0.113	0.155

Notes: The table reports the tests for differential attrition by monitoring and gender treatment assignments. *Main Sample* is a binary indicator equal to 1 if the evaluation of a given CV is included in the main sample. An evaluation is not included in the main sample if all the preregistered attention checks are not met: the respondent understood the IRR exercise, and a given block of eight profiles has some variation in the evaluations. The dependent variables are a binary indicator equal to 1 if the respondent is randomly assigned to: Monitoring Behavior (MB) arm in column (1); Monitoring Safety (MS) arm in column (2), and Female profile in column (3). Samples used to estimate the regressions are restricted to Pure Control (PC) and MB, PC and MS, and PC only in columns (1)-(3), respectively. Standard errors in brackets are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Experimental Results – OLS Gender by Ability Index in Pure-Control Group

			Stated preferences for Workforce Gender Mix				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meet (0-1)	Main Sample	Gender-Mix Pref. > 0	Bottom QT	2nd QT	3rd QT	4th QT	Top QT
Female	-0.104***	-0.078***	-0.231***	-0.138***	-0.100***	-0.090***	0.026
	[0.016]	[0.016]	[0.050]	[0.039]	[0.025]	[0.026]	[0.033]
Female x Ability Index	-0.009	-0.010	-0.010	-0.029	0.021	-0.043*	0.000
	[0.011]	[0.012]	[0.010]	[0.032]	[0.017]	[0.024]	[0.030]
Ability Index	0.133***	0.131***	0.141***	0.170***	0.104***	0.167***	0.107**
	[0.017]	[0.018]	[0.044]	[0.053]	[0.033]	[0.038]	[0.046]
Observations	7,365	6,206	1,159	894	2,340	1,740	1,232
R-squared	0.133	0.133	0.212	0.195	0.157	0.146	0.144
Control mean	0.478	0.480	0.466	0.460	0.503	0.477	0.451

Notes: The table reports the treatment effects from the estimation of the regression model 3 for the sample of employers in the Pure-Control group (PC), and by quintiles distribution of employers' stated preferences for gender workforce composition. Column (1) shows the effects from the main sample; column (2) excludes employers with stated preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintiles distribution of stated preferences for gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a female candidate. *Ability Index* is a standardized index that combines all high-quality trainee characteristics in the CVs with equal weights: completion of secondary education, a Directorate of Industrial Training (DIT) certification, a 12- (vs. 6-month) study period, and a high GPA. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Experimental Results – OLS Gender by Ability in Monitoring Arms

Panel A: **Monitor Behavior**

			Stated Preferences for Workforce Gender Mix				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meet (0-1)	Main Sample	Gender-Mix Pref. > 0	Bottom QT	2nd QT	3rd QT	4th QT	Top QT
Female	-0.153***	-0.148***	-0.169***	-0.249***	-0.147***	-0.094**	-0.140***
	[0.024]	[0.025]	[0.053]	[0.064]	[0.036]	[0.038]	[0.050]
Female x High Ability	-0.028	-0.015	-0.117**	0.077	-0.021	-0.063	-0.021
	[0.023]	[0.023]	[0.048]	[0.057]	[0.037]	[0.048]	[0.054]
High Ability	0.111***	0.107***	0.147***	0.010	0.099***	0.190***	0.103**
	[0.021]	[0.020]	[0.040]	[0.044]	[0.032]	[0.037]	[0.046]
Observations	6,830	5,847	983	1,070	2,134	1,620	1,023
R-squared	0.148	0.150	0.205	0.190	0.185	0.132	0.182
Control mean	0.516	0.523	0.475	0.555	0.510	0.499	0.556

Panel B: **Monitor Safety**

			Stated preferences for Workforce Gender Mix				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meet (0-1)	Main Sample	Gender-Mix Pref. > 0	Bottom QT	2nd QT	3rd QT	4th QT	Top QT
Female	-0.005	0.027	-0.228**	-0.029	0.000	0.046	0.109
	[0.025]	[0.025]	[0.091]	[0.042]	[0.040]	[0.041]	[0.071]
Female x High Ability	-0.024	-0.029	-0.011	0.008	0.005	-0.084**	-0.019
	[0.025]	[0.027]	[0.047]	[0.037]	[0.043]	[0.040]	[0.068]
High Ability	0.081***	0.090***	0.058	0.017	0.084*	0.131***	0.070
	[0.016]	[0.016]	[0.047]	[0.023]	[0.044]	[0.022]	[0.055]
Observations	6,530	5,695	835	1,062	1,997	1,860	776
R-squared	0.102	0.103	0.219	0.178	0.111	0.116	0.103
Control mean	0.468	0.452	0.556	0.437	0.454	0.446	0.484

Notes: The table reports the treatment effects from estimation regression models 3 for the sample of respondents in the Monitoring Behavior (MB) arm (Panel A) and in the Monitoring Safety (MS) arm (Panel B), and by quintiles distribution of preferred gender workforce composition of the employer. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintiles distribution of preferred gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants to meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a female candidate. *High-Ability* is a dummy indicating whether the résumé has a top GPA score (5/5) within their vocational training program. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Experimental Results – OLS Gender by Ability Index in Monitoring Arms

Panel A: **Monitor Behavior**

			Stated Preferences for Workforce Gender Mix				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meet (0-1)	Main Sample	Gender-Mix Pref. > 0	Bottom QT	2nd QT	3rd QT	4th QT	Top QT
Female	-0.168***	-0.155***	-0.228***	-0.210***	-0.158***	-0.126***	-0.150***
	[0.017]	[0.018]	[0.051]	[0.050]	[0.028]	[0.033]	[0.038]
Female x Ability Index	0.001	0.010	-0.038*	0.022	0.002	-0.019	0.045
	[0.013]	[0.014]	[0.022]	[0.038]	[0.024]	[0.019]	[0.029]
Ability Index	0.121***	0.121***	0.128***	0.050	0.110***	0.210***	0.094
	[0.020]	[0.021]	[0.045]	[0.048]	[0.027]	[0.042]	[0.057]
Observations	6,830	5,847	983	1,070	2,134	1,620	1,023
R-squared	0.148	0.150	0.204	0.189	0.185	0.131	0.184
Control mean	0.538	0.545	0.500	0.540	0.502	0.539	0.648

Panel B: **Monitor Safety**

			Stated preferences for Workforce Gender Mix				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meet (0-1)	Main Sample	Gender-Mix Pref. > 0	Bottom QT	2nd QT	3rd QT	4th QT	Top QT
Female	-0.017	0.013	-0.234***	-0.025	0.002	0.004	0.099
	[0.022]	[0.021]	[0.079]	[0.038]	[0.035]	[0.035]	[0.060]
Female x Ability Index	-0.017	-0.019	-0.042	-0.009	-0.004	-0.027	-0.027
	[0.012]	[0.013]	[0.030]	[0.018]	[0.023]	[0.023]	[0.033]
Ability Index	0.095***	0.104***	0.085	0.030	0.110***	0.125***	0.090
	[0.016]	[0.015]	[0.058]	[0.035]	[0.034]	[0.026]	[0.063]
Observations	6,530	5,695	835	1,062	1,997	1,860	776
R-squared	0.102	0.104	0.221	0.178	0.111	0.115	0.104
Control mean	0.451	0.447	0.480	0.382	0.458	0.458	0.485

Notes: The table reports the treatment effects from estimation regression models 3 for the sample of respondents in the Monitoring Behavior (MB) arm (Panel A) and in the Monitoring Safety (MS) arm (Panel B), and by quintiles distribution of preferred gender workforce composition of the employer. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintiles distribution of preferred gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants to meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a female candidate. *Ability Index* is a standardized index that combines all high-quality trainee characteristics in the CVs with equal weights: completion of secondary education, a Directorate of Industrial Training (DIT) certification, a 12- (vs. 6-month) study period, and a high GPA. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: OLS Gender by Monitoring – Secondary Outcomes

	(1) Quality	(2) Trust-worthiness	(3) Effort	(4) Expected Earnings	(5) Net Pay
Female	-0.169***	0.055**	-0.208***	-12.390***	5.307
	[0.026]	[0.026]	[0.028]	[1.595]	[3.881]
Monitor Behavior	-0.013	-0.033	-0.011	2.405	-2.214
	[0.041]	[0.049]	[0.042]	[3.198]	[5.507]
Monitor Safety	-0.029	-0.034	-0.014	-1.473	2.688
	[0.048]	[0.053]	[0.044]	[3.204]	[5.560]
Female x Monitor Behavior	-0.042	-0.009	-0.055	-0.940	1.732
	[0.028]	[0.035]	[0.037]	[2.417]	[3.871]
Female x Monitor Safety	0.034	0.062	0.018	-0.557	-12.153*
	[0.047]	[0.043]	[0.044]	[2.430]	[6.164]
Observations	20,725	20,614	20,685	20,726	7,025
R-squared	0.226	0.083	0.173	0.195	0.135
PC Mean	0.086	-0.026	0.106	114.767	-25.461

Notes: The table reports the treatment effects for the main sample, by pre-registered secondary outcomes. Columns (1)-(3) reports standardized outcomes relative to PC group. Column (1) shows estimates for the perceived quality and skills of candidates: "Based on your first impression, 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.". Column (2) shows estimates for the perceived trustworthiness of candidates: "Based on your first impression, how would you rate the workers' behavior (trustworthiness and honesty)? Please rate on 0 to 10, where 0 is not at all trustworthy/honest and 10 is very trustworthy/honest.". Column (3) shows estimates for the perceived effort of candidates: "Based on your first impression, how likely is this worker to put in effort (hardworking, not lazy, concentrated, punctual)? Please rate on a scale from 0 to 10, where 0 is very unlikely and 10 is very likely.". Column (4) shows estimates for the expected monthly earnings of the candidates in USD: "What is your best guess of the monthly earnings of this worker a year from now?". Column (5) shows estimates for the net monthly pay in USD for candidates during the probation period, and it's calculated as the monthly wage less the fees the trainee is asked to pay. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: OLS Monitoring — Meet

	(1) Meet
Monitor Behavior	0.019*
	[0.011]
Monitor Safety	0.030**
	[0.011]
Observations	20,725
R-squared	0.114
Control Mean	0.470

Notes: The table reports the treatment effects for the main sample. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate for a probation period at the firm. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. The control mean is that of the *Pure Control* arm. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: OLS Gender by Monitoring – Resume Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Meet (0-1)	Age ≥ Median	Married	Second. Educ.	English	Trained 1+ year	DIT Certified
CHRS	-0.094***	-0.209***	0.007	0.067***	0.209***	0.145***
	[0.012]	[0.011]	[0.012]	[0.018]	[0.011]	[0.013]
Monitor Behavior	0.012	0.009	0.025*	0.016*	0.029**	0.029***
	[0.014]	[0.012]	[0.015]	[0.009]	[0.012]	[0.010]
Monitor Safety	0.037***	0.022*	0.034**	0.033***	0.040***	0.038***
	[0.014]	[0.012]	[0.015]	[0.009]	[0.012]	[0.010]
CHRS x Monitor Behavior	0.010	0.019	-0.010	0.016	-0.019	-0.035*
	[0.017]	[0.016]	[0.018]	[0.025]	[0.016]	[0.018]
CHRS x Monitor Safety	-0.008	0.019	-0.003	-0.009	-0.019	-0.021
	[0.017]	[0.017]	[0.018]	[0.026]	[0.017]	[0.018]
Observations	20,725	20,725	20,725	20,725	20,725	20,725
Control Mean	0.470	0.574	0.466	0.462	0.365	0.428

Notes: The table reports the treatment effects for the main sample, by resume characteristics (CHRS). Estimates are shown in Columns (1)–(6), respectively, for whether the candidate: is above the median age (22.5 years), is married, has attended secondary education, speaks English, have been attending VTI one- or two-year training relative to 6-month training, holds a Directorate of Industrial Training (DIT) certification. Strata fixed effects are included (not reported in the table). Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: OLS Gender by Monitoring – Heterogeneous Treatment Effects

	Gender-Mix Pref. > 0		Employer Gender		VTI Search	
Meet (0-1)	(1) Low	(2) High	(3) Female	(4) Male	(5) Yes	(6) No
Female	-0.146***	-0.041*	0.001	-0.105***	-0.086***	-0.174***
	[0.021]	[0.022]	[0.111]	[0.017]	[0.018]	[0.042]
Monitor Behavior	0.035*	0.080***	0.149	0.049***	0.067***	-0.027
	[0.019]	[0.020]	[0.100]	[0.014]	[0.017]	[0.032]
Monitor Safety	-0.021	-0.006	0.280***	-0.018	-0.014	-0.008
	[0.018]	[0.024]	[0.075]	[0.016]	[0.017]	[0.041]
Female x Monitor Behavior	-0.043	-0.092***	-0.206	-0.061**	-0.086***	0.027
	[0.031]	[0.032]	[0.138]	[0.023]	[0.024]	[0.065]
Female x Monitor Safety	0.098***	0.075*	-0.218*	0.092***	0.081**	0.096
	[0.033]	[0.041]	[0.125]	[0.028]	[0.032]	[0.069]
Observations	12,450	8,275	727	19,998	17,108	3,545
R-squared	0.147	0.112	0.173	0.130	0.127	0.172
Control Mean	0.528	0.520	0.465	0.524	0.513	0.566

Notes: The table reports the treatment effects for the main sample, by pre-registered heterogeneity dimensions: Columns (1)-(2) show estimates for above- (high) and below-median (low) optimal gender composition of the workforce; Columns (3)-(4) show estimates for whether employers typically hire from vocational training institutes or not; Columns (5)-(6) split the sample by the employers' gender. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants to meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a woman. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: OLS Gender by Ability and by Monitoring — Meet

	(1) Meet
Female	-0.087***
	[0.022]
High Ability	0.118***
	[0.018]
Female x High Ability	-0.033
	[0.032]
Monitor Behavior	0.055***
	[0.018]
Monitor Safety	0.006
	[0.019]
Female x Monitor Behavior	-0.067**
	[0.027]
Female x Monitor Safety	0.082**
	[0.030]
High Ability x Monitor Behavior	-0.007
	[0.021]
High Ability x Monitor Safety	-0.038*
	[0.022]
Female x High Ability x Monitor Behavior	0.005
	[0.030]
Female x High Ability x Monitor Safety	0.009
	[0.029]
Observations	20,726
R-squared	0.126
Mean: Male, Low Ability in PC	0.464

Notes: The table reports the treatment effects for the main sample. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a woman. *High-Ability* is a dummy indicating whether the résumé has a top GPA score (5/5) within their vocational training program. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. The control mean is that of Male candidates with Low Ability (i.e. median GPA score within their program) in *Pure Control* arm. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: OLS Gender by Monitoring and Lin. Interaction with Gender Pref.

Meet (0-1)	(1) Pure Control	(2) Monitor Behavior	(3) Monitor Safety	(4) Main Sample
Female	-0.219***	-0.217***	-0.139***	-0.218***
	[0.034]	[0.033]	[0.044]	[0.034]
Monitor Behavior				-0.002
				[0.027]
Monitor Safety				-0.026
				[0.029]
Gender-Mix	-0.106	0.132*	-0.048	-0.111
	[0.084]	[0.077]	[0.101]	[0.080]
Female x Monitor Behavior				0.002
				[0.046]
Female x Monitor Safety				0.076
				[0.054]
Female x Gender-Mix	0.521***	0.224**	0.564***	0.514***
	[0.121]	[0.105]	[0.165]	[0.122]
Monitor Behavior x Gender-Mix				0.243**
				[0.101]
Monitor Safety x Gender-Mix				0.062
				[0.124]
Female x Monitor Behavior x Gender-Mix				-0.296*
				[0.161]
Female x Monitor Safety x Gender-Mix				0.062
				[0.209]
Observations	7,365	6,830	6,530	20,725
R-squared	0.130	0.144	0.105	0.134
Baseline Mean	0.522	0.572	0.522	0.552

Note: The table reports the treatment effects on the interest of employers to meet the candidates by stated preferences for workforce gender mix of employers. Columns (1)-(3) include all respondents assigned to the Pure Control (PC), Monitoring Behavior (MB), and Monitoring Safety (MS) arms in the between-subject design (main specification). Column (4) includes all respondents in the main sample and estimates a fully interacted model with candidate gender, monitoring treatment, and employer stated preferences for workforce gender mix. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants to meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the resume corresponds to a woman. *Monitoring Behavior* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a binary indicator equal to 1 if the firm was randomly assigned to audit visits to monitor trainees' safety. *Gender-Mix* is a variable ranging from 0 to 1 that measures the share of female workers preferred by the employer in their ideal workforce composition. *Baseline Mean* is the average of *Meet* for male resumes evaluation by employers with stated preferences for an all-male workforce in PC. Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered at the employer and profile levels (reported in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

Table A12: OLS Gender by Monitoring Arm for Within-subject Design

			Stated Preferences for Workforce Gender Mix				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Meet (0-1)	Main Sample	Gender-Mix Pref. > 0	Bottom QT	2nd QT	3rd QT	4th QT	Top QT
Female	-0.115***	-0.096***	-0.236***	-0.122**	-0.100***	-0.122***	-0.050**
	[0.015]	[0.015]	[0.055]	[0.049]	[0.024]	[0.026]	[0.024]
Monitor Behavior	0.092***	0.103***	0.037	0.076**	0.121***	0.083***	0.106***
	[0.013]	[0.013]	[0.038]	[0.032]	[0.017]	[0.025]	[0.031]
Monitor Safety	-0.046***	-0.045***	-0.018	-0.092***	-0.041*	-0.050**	-0.027
	[0.013]	[0.014]	[0.043]	[0.027]	[0.024]	[0.021]	[0.025]
Female x Monitor Behavior	-0.122***	-0.125***	-0.119*	-0.149**	-0.129***	-0.096**	-0.123***
	[0.020]	[0.020]	[0.064]	[0.059]	[0.029]	[0.039]	[0.036]
Female x Monitor Safety	0.149***	0.163***	-0.019	0.179***	0.163***	0.174***	0.166***
	[0.024]	[0.024]	[0.085]	[0.057]	[0.042]	[0.040]	[0.046]
Observations	20,800	18,409	2,391	2,447	6,521	5,714	3,727
R-squared	0.221	0.214	0.314	0.249	0.237	0.208	0.191
Control Mean	0.529	0.529	0.536	0.522	0.513	0.550	0.526
p-value MB vs. MS	0.000	0.000	0.240	0.000	0.000	0.000	0.000

Note: The table reports the treatment effects for the sample for which we implemented the within-subject randomization. The table also reports the treatment effects by quintile distribution of preferred gender workforce composition of the employer. Column (1) shows the effects from the main sample; column (2) excludes employers with preferences for an all-male workforce. Columns (3) to (7) restrict the sample to each quintile's distribution of preferred gender workforce composition. The dependent variable *Meet* is a binary indicator equal to 1 if the employer wants the meet the candidate for a probation period at the firm. *Female* is a binary variable indicating whether the résumé corresponds to a woman. *Monitoring Behavior* is a dummy indicating whether the firm was randomly assigned to audit visits to monitor trainees' behavior. *Monitoring Safety* is a dummy indicating whether the firm was randomly assigned to audit visits to monitor trainees' safety. We report results for the primary outcome variable: interest of the employer to meet (0-1). Respondent and profile-by-ability fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile-by-ability levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: OLS Gender by Monitoring – Offer

	(1) Offer | Meet
Female	-0.023**
	[0.009]
Monitor Behavior	-0.030**
	[0.015]
Monitor Safety	-0.029**
	[0.013]
Female x Monitor Behavior	-0.007
	[0.014]
Female x Monitor Safety	0.003
	[0.009]
Observations	3,424
R-squared	0.172
Control Mean	0.621

Notes: The table reports treatment effects for the subsample of profile evaluations selected by respondents, for which data on the outcome variable *Offer* is available:"How likely would you be to offer this worker a position as a mechanic at your firm? Please rate on a scale from 0 to 10, where 0 is very unlikely and 10 is very likely.". Profile and strata fixed effects are included (not reported in the table). Standard errors are clustered by respondent and profile levels. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Trainee Outcomes and Employer's Beliefs

Panel A: **Trainees**

	Social Preferences				Skills		Supervision & Honesty Game				Dictator Game	Trust Game		Cooperation Game
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Patience	Altruism	Pos. Recipr.	Trusting	Theory	Practice	Task Completed Unsupervised	Task Completed Supervised	Lying Unsupervised	Lying Supervised	Share	Trust Pl. 1	Reciprocate Pl. 2	Cooperate
Female	0.097	0.287*	0.370**	-0.049	0.105	0.009	-0.040**	-0.058***	0.086**	0.054	-0.137	-0.250***	-0.075	-0.045
	[0.145]	[0.144]	[0.101]	[0.092]	[0.082]	[0.040]	[0.014]	[0.013]	[0.033]	[0.028]	[0.111]	[0.045]	[0.054]	[0.059]
Observations	171	171	171	171	187	182	182	182	182	182	182	182	182	182
R-squared	0.002	0.020	0.039	0.001	0.016	0.000	0.023	0.042	0.019	0.018	0.020	0.061	0.010	0.003
Control Male	0.000	0.000	0.000	0.000	0.752	0.920	0.989	0.989	0.071	0.018	0.723	0.679	0.875	0.830

Panel B: **Employers**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Female	-0.061	-0.084*	0.162*	0.481***	-0.121**	-0.154**	0.047*	0.000	-0.115***	-0.057***	-0.089**	-0.011	-0.035*	-0.018**
	[0.054]	[0.024]	[0.048]	[0.004]	[0.020]	[0.020]	[0.015]	[0.005]	[0.011]	[0.003]	[0.011]	[0.005]	[0.009]	[0.002]
Observations	919	919	920	921	1,205	1,206	1,842	1,829	1,831	1,826	1,838	1,835	1,829	1,831
R-squared	0.005	0.009	0.039	0.312	0.048	0.067	0.013	0.000	0.071	0.019	0.030	0.001	0.006	0.002
Control Male	0.435	0.449	0.253	0.130	0.580	0.607	0.646	0.881	0.466	0.256	0.573	0.534	0.602	0.571

Notes: The table reports gender differences in trainees' outcomes (Panel A) and the corresponding employer beliefs (Panel B). Columns (1)–(4) present standardized survey measures of social preferences following Falk et al. (2018): patience, altruism, positive reciprocity, and trust, respectively. In Panel A, the coefficient on *Female* captures the difference between female and male trainees. In Panel B, *Female* captures the difference in employer beliefs about female versus male trainees. Columns (5)–(6) present the share of correct answers to sector-specific theory and practical skill questions, respectively. Theory questions are randomly drawn from the technical exam taken by trainees prior to the survey. Practical skills are measured through a self-reported binary indicator on the ability to perform a specific technical task, designed by VTI instructors. The remaining columns shows results of four behavioral games. Columns (7)–(8) report the share of tasks completed in a novel lab-style task — Supervision & Honesty Game — by trainees (Panel A), and employers' beliefs about the same task outcomes under conditions without and with supervision (Panel B). Columns (9)–(10) report the share of lying behavior in the same game. Column (11) reports sharing behavior in a dictator game. Columns (12)–(13) correspond to player 1 (trusting) and player 2 (reciprocating) behavior in a trust game. Column (14) reports decisions to cooperate in a cooperation game. Standard errors are clustered at VTI level in Panel A and at the sector level in Panel B. Standard errors in brackets * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.