IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14083

Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs

Alexia Delfino

JANUARY 2021

DISCUSSION PAPER SERIES

IZA DP No. 14083

Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs

Alexia Delfino
Bocconi University and IZA

JANUARY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs*

Traditionally female-dominated sectors are growing and male-dominated ones shrinking, yet sectorial male shares are not changing. Why? I embed a field experiment within the UK national recruitment program for social workers to analyse barriers to men's entry and the nature of men's sorting into female-dominated occupations. I modify the content of recruitment messages to potential applicants to exogenously vary two key drivers of selection: perceived gender shares and expectations of returns to ability. I find that perceived gender shares do not affect men's applications, while increasing expected returns to ability encourages men to apply and improves the average quality of the applicants. This allows the employer to hire more talented men, who consistently perform better on the job and are not more likely to leave vis-á-vis men with lower expected returns to ability. I conclude by showing that there is no trade-off between men's entry and women's exit among talented applicants, both at hiring and on-the-job, and thus the net impact of raising expected returns to ability for the employer is positive.

JEL Classification: D23, D83, J24, J7, M5
Keywords: recruitment experiment, gender barriers, beliefs

Corresponding author:
Alexia Delfino
Bocconi University
Via Roberto Sarfatti, 25
20100 Milano MI
Italy
E-mail: alexia.delfino@unibocconi.it

* I thank the Director of Selection and all the members of the Attraction and Selection teams in the partner organization. I thank Nava Ashraf, Oriana Bandiera, Matthew Levy and Erik Eyster for their priceless advice throughout the project. I also thank advice from Robin Burgess, Maitreesh Ghatak, Gharad Bryan, Edward Glaeser, Rachel Kranton, Marco Manacorda, Eliana La Ferrara, Thomas Le Barbanchon, Robert Akerlof, Steve Pischke, Florian Englmaier, Matthew Lowe, Johanna Rickne, Paul Gertler and Rocco Macchiavello. Francesco Sannino, Matteo Benetton, Michel Azulai, Nicola Limodio, Giulia Giupponi and Clement Minaudier provided helpful comments. I thank Miguel Espinosa for contributing to multiple rounds of revision. I thank Cristhian Acosta, Alex Blums, Andres Cordoba and Pedro Cabra for excellent research assistance. I acknowledge funding from ESRC and STICERD. A special acknowledgement to focus group participants who gave invaluable inputs on the design. The experiment is registered in the AEA RCT Registry with ID AEARCTR-0002351 and was approved by the LSE Ethics Committee in August 2017.

1 Introduction

In the last half century, the shift from brawn-intensive to brain-intensive occupations has eroded the traditional advantage that men enjoyed in the labour market. The manufacturing share of employment in the US fell from 29.7 to 12.7 percent between 1968 and 2008, while the service share rose from 56 to 75 percent in the same period (Ngai and Petrongolo, 2017). Other OECD countries share similar trends, which increase the risk of involuntary displacement and long spells of male unemployment (OECD, 2019). Female-dominated industries, such as health and education, displayed the highest growth, and yet their gender composition barely changed despite falling male labour force participation (Blau and Kahn, 2017). Understanding the barriers to men's entry in these occupations is important to help workers in declining industries move towards new opportunities.

These trends motivate my research question: why don't men enter female-dominated jobs? Providing an answer requires two steps. First, we need to know whether men do not apply. I generate exogenous variation in the recruitment messages to potential applicants for a job in social work, a fast-growing female-dominated occupation, to back out supply-side barriers that hinder men's entry into female-dominated professions. However, more male applicants does not necessarily lead to more male workers, because employers will not hire untalented applicants. We thus need to know whether firms can hire more men without lowering the quality of their workforce. This second step requires following what happens to applicants of both genders beyond the hiring process.

I design and run a large-scale natural field experiment (List and Rasul, 2011; List and Metcalfe, 2014) to bring into a controlled-setting a real-world policy that aims at increasing gender diversity in female-dominated jobs. I embed the experiment within the UK-wide recruitment program for social workers, where I observe applications as well as hiring and on-the-job outcomes over two years for candidates of both genders. This allows me to say whether - and how - bringing more men into female-dominated jobs is good for employers and whether this has spillovers on women's selection.

The design aims to disentangle the role of two factors commonly believed to influence the attractiveness of female-dominated jobs for men: perceived gender shares and expected returns to ability. The former embodies occupational features related to the association between an occupation and a certain gender, which have been shown to affect labour supply (Akerlof and Kranton, 2000; Flory et al., 2015; Del Carpio and Guadalupe, 2018; Oh, 2019). For instance, a man might fear going against his male identity by entering an occupation with few men. The expected returns channel addresses informational constraints about the potential for being effective on the job and related rewards (Reuben et al., 2017; Wiswall and Zafar, 2018; Samek, 2019). While there is a rich stream of work on the role of beliefs in affecting discriminatory behavior (Hoff and Pandey, 2006; Bordalo et al., 2016; Glover et al., 2017; Bohren et al., 2019; Coffman et al., 2019), research on how beliefs about self-efficacy affect individual self-selection into jobs is scant. Men may not know whether jobs to which they have little exposure, such as female-dominated ones, offer them opportunities to be successful. Success and recognition have traditionally been important determinants of men's work satisfaction (Goldin, 2006) and seeing only a few highly-selected members of their own gender may create uncertainty about the possibility of getting rewards for talent in female-jobs.[1]

[1] People might care about rewards to ability for extrinsic reasons, if performance is tied to incentives or career promotions,

Recruitment ads are an increasingly common business tool to increase diversity (Flory et al., 2019; Perkins et al., 2000; Tipper, 2004), but unpacking their effect on the demographic composition and the talent of the pool of applicants is challenging. First, independent sources of exogenous variation are required to disentangle the effect of different contents. Second, participants' outcomes beyond applications are needed to assess the talent of the recruits and the net benefit for the employer. A controlled setting with long-term outcomes allows me to overcome these challenges.

Furthermore, I am able to minimize the risk that the intervention may artificially affect the behaviour of the participants through experimenter demand effects (Levitt and List, 2007) by embedding the experiment within a real organization and overlaying the randomized test of new recruitment messages on potential applicants in their natural environment. Subjects are not aware of participating in an experiment and, at the same time, the employer does not know the mapping between candidates and treatment assignment. By conducting the experiment in such a double-blind manner, I do not interfere with the natural course of the hiring process and I can follow participants from applications to job offers and, afterwards, on the job, without undue influence from researchers. Compared to alternative sources of variation, for instance monetary incentives, my design preserves the organizational systems in place, does not require hands-on administration and mimics common low-cost policies that employers use to increase gender diversity. These features make the design easy to scale and replicate in other contexts (Al-Ubaydli et al., 2017).[2]

To generate exogenous variation in perceived gender shares in the job, I included photographs of real workers in the job advertisement and randomized whether the portrayed worker was of the same or different gender to the potential applicant (Bertrand and Mullainathan, 2004; Benjamin et al., 2010). Auxiliary survey evidence that I gathered shows that the photographs have the intended impact of creating a wedge in perceived gender shares (6 percentage points on average, 9% of the average female share).Different channels similarly predict that a male photograph achieves a positive utility shock for men by increasing the perceived male share in a female-job. For instance, working in a female-dominated job might threaten men's identity (Akerlof and Kranton, 2000, 2005; Del Carpio and Guadalupe, 2018) or internalized social norms (Bursztyn and Jensen, 2017; Bursztyn et al., 2020; Baranov et al., 2020). Men may also have an innate distaste for working with a majority of women (Becker, 1957) or anticipate employers', clients' or coworkers' preferences for female workers (Folke and Rickne, 2020).[3]

To shock expected returns to ability, I reported in the job advert the aggregate performance of a selected past cohort of workers from the organization, who had either moderate or high success.[4] Half of the sample were informed that, in a previous year, 66% of workers were high-performers and the other half that 89% of workers were high-performers. I interpret these statistics as indicating high and low

or intrinsic motivation, if they care about social recognition, feeling competent or about the actual impact generated in the job.

[2] The organization does not use performance bonuses or other monetary incentives. The effect of introducing them for the first time could create novelty or surprise effects on the participants, which would confound the interpretation of the experiment as a change to expected returns to ability.

[3] Some authors find evidence of employer's discrimination against men in female-dominated jobs (Booth and Leigh, 2010; Rich, 2014). This explanation is second order in my context, where the employer wants to attract more men and trains its recruiters against implicit biases in the selection of men and minorities (Bertrand et al., 2005).

[4] I use actual records of the organization in the previous three years. This allowed the communication of truthful but partial information, which on average affects beliefs differently between experimental groups (Dal Bó et al., 2017).

marginal returns to ability on the job, respectively. Intuitively, lower past success (66%) signals a more challenging environment where individual ability can make a large difference. Auxiliary survey data I collected confirm that the two statistics create a wedge in expectations of success, but differentially for high and low ability people, as one should expect when returns to ability are different.[5] This second variation captures the fact that being in a minority in a job imposes informational constraints on men, which interact with their job-specific talent. The partner organization was indeed worried that talented men would not apply because they thought social work was an "easy job". We co-designed the experimental treatments to address these perceptions.

The first result of the paper is that perceived gender shares do not affect men's applications. This is surprising in light of many policy proposals trying to attract men to teaching or nursing by portraying more male nurses or teachers in job adverts (The Economist, 2018; Flynn, 2006; McGonagle, 2019). For instance, portraying more male nurses is one of the pillars of the biggest recruitment drive in the history of the UK National Health System (NHS from here on, see McGonagle, 2019), but my results suggest that the gender composition of portrayed actors might not matter as much as was thought.[6]

In contrast, expectations of higher returns to ability increase men's applications by 15% vis-à-vis expectations of lower returns. This means that being informed about moderate past performance (as opposed to outstanding) encourages men to apply more. This second result is novel and contrasts with most role model interventions to encourage women in STEM, whose standard design provides participants with examples of high past success (Porter and Serra, 2020; Breda et al., 2020; Del Carpio and Guadalupe, 2018). My paper shows that information about high success can be interpreted as a signal of low returns to ability rather than the unconditional probability of success, which might encourage only low-ability people to apply.

The magnitude of the effect of increasing expected returns to ability is large, which is noteworthy given that the treatment is cost-free for the employer. For a candidate with average job-specific ability, the estimated effect is comparable to an 11% increase in the wage, which would cost the employer more than a million dollars a year to implement. How can information have such a big impact? According to Bayesian updating, the magnitude of the effect of information is proportional to men's uncertainty in priors. Thus the substantial gap in application rates between the two information arms suggests that informational constraints are an important barrier to men's entry in female-dominated jobs. This interpretation is confirmed by heterogeneous treatment effects, as I find that the difference in application rates between the two information arms is larger among men coming from gender-segregated labour markets (Jensen, 2010; Charles et al., 2018; Baranov et al., 2020).

The third result of the paper is that raising expected returns to ability allows the employer to hire and retain more talented male workers. Applicants with higher expected returns to ability are better in terms of observable characteristics - such as academic achievement - and consequently receive more job offers. Crucially, in the first two years on the job, men attracted by information on higher returns to ability show a consistently higher performance than those who received the low returns to ability treatment (by 0.25 SD), and they are no more likely to leave the job and report higher levels of job

[5] High ability people are more likely to believe that they are better than the median applicant when they receive information on high than low expected returns to ability, while the opposite happens for low ability people.

[6] However, this null result is consistent with other studies which find little room for occupation-specific preferences and gender-shares in explaining career choices (Wiswall and Zafar, 2018; Hsieh et al., 2019).

satisfaction and intent to stay. This is particularly important in a sector with high levels of burnout and where 50% of workers plan to stay less than two years (Ravalier, 2019).

Interpreted through the lens of a Roy-type framework, my results suggest that men are negatively sorted in the job and that the marginal male applicant faces an outside option which has steeper returns to ability than the average applicant.[7] In the model, an increase in expected returns to ability increases the gradient of utility with respect to job-specific talent. This rotation increases utility for high ability applicants and decreases utility for low ability people. The prediction is that changing expected returns to ability may improve the quality of applicants through the entry of more talented applicants (if the least talented men usually sort into social work) or the exit of less talented ones (if the most talented men usually sort into social work). My results are consistent with the former case: increasing expected returns to ability in the job attracts not only more men, but also improves the quality of male applicants.

However, the net benefit of raising expected returns to ability for the employer is unclear if we do not quantify spillovers on the selection of women, who still represent the majority in the job. I find that women are insensitive to information provision on average, which could lead to the conclusion that changing expectations of returns to ability is a silver bullet for the employer: it achieves higher diversity, better quality among the gender minority and has no effect on the majority. Nevertheless, the very fact that men start entering social work might have an impact on women's behavior. If male shares in female-dominated jobs increased in the long run, what would be the impact on women's choices? Using the variation in perceived male shares generated by the photograph manipulation, I find that women apply less when they believe that there are more male social workers in the job (a difference of 7.5% between the male and female photograph arm), but this has no impact on their average quality. At the same time, low-performing female recipients of the male photograph are more likely to quit the job. This latter result creates a trade-off for the employer: on the one hand female stayers are better, but on the other hand turnover is larger. I ran a survey with the recruitment personnel of the partner organization and found that the majority of them deem the resulting increase in performance to be high enough to justify the higher turnover.

Taken together, my results suggest that breaking informational barriers to men's entry in female-dominated jobs might increase gender diversity and improve overall workforce quality in a gender-neutral way. This yields an optimistic message for policy. Both the stigma associated with working in a female occupation and men's perceptions of their returns to ability have been central in the debate around the conversion of unemployed men into service jobs. The two have different policy implications.[8] The femaleness associated with some occupations may be difficult to modify and changes in gender composition take time. While people can be monetarily compensated or compositional changes can be accelerated through quotas, uncertain or incorrect expectations can be more cheaply tackled through information provision and incentives, for example through low-cost organizational practices that recognize good performance.

This paper contributes to the growing empirical work on social identity (Del Carpio and Guadalupe,

[7] Empirically, I find that the increase in men's applications in the high expected returns to ability treatment is indeed driven by men who face a higher-than-median wage dispersion in the UK labour market. See Section 7.1.

[8] See, for instance, this article (Miller, 2017): https://www.nytimes.com/2017/01/04/upshot/why-men-dont-want-the-jobs-done-mostly-by-women.html.

2018) and stereotypes (Steele and Aronson, 1995; Stone et al., 1997; Hoff and Pandey, 2006; Bordalo et al., 2016; Glover et al., 2017; Alan et al., 2018; Carlana, 2019) by disentangling the effect of group composition from limited information in men's and women's identity-related choices. Relatedly, while studies about gender differences across male-typical or female-typical domains abound in the lab (Bordalo et al., 2019), this is the first paper to study the determinants of men's choices in a real field setting. The paper also speaks to the personnel economics literature on the effects of posted wages or amenities on the applicant pool (Dal Bó et al., 2013; Marinescu and Wolthoff, 2020; Ashraf et al., 2020; Deserranno, 2019; Abebe et al., 2017; Flory et al., 2019), by showing that posted adverts might address information frictions which prevent minorities from applying for jobs which are uncommon for their demographics. Finally, my paper shows the relevance of subjective expectations (Nguyen, 2008; Jensen, 2010; Zafar, 2013; Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015; Reuben et al., 2017; Boneva et al., 2017; Wiswall and Zafar, 2018) in an understudied, but highly relevant empirical setting (Katz, 2014).

2 Institutional context

There are several reasons why social work is a good setting for my research questions. Women historically represent more than 70% of social workers in the US and in the UK (Figure A.1a). In a similar way to many other female-dominated service occupations, the demand for social workers is expected to grow (Figure A.1b) as this is a sector that faces looming staff shortages in the following decades (Lin et al., 2015). For instance, the growth rate of social workers is expected to be twice the average growth across all US occupations, and to be greater in areas of high male joblessness (see Figure A.1c).

During 2017, I collaborated with one of the main UK recruiters of public sector social workers. The organization offers a two-year on-the-job training position targeted at either final year students from all disciplines or current workers from any industry. Previous experience in social work is not required, a feature which allows me to study entry into this occupation broadly, not only looking at people with job-specific training. Moreover, this recruitment strategy implies that the sample is heterogeneous in terms of background exposure to social work, making informational and psychological constraints particularly relevant for some. Every year, new hires are assigned to teams allocated to Local Authorities across England and earn a stipend which is comparable to the average UK annual entry salary in social services (26k GBP), primary school teaching (24k GBP) and nursing (22k GBP). The daily job involves both office tasks (e.g., case writing) and meetings with families in need and other stakeholders, such as lawyers, medical professionals and the police. After the programme, the majority of workers stay in similar positions (between 60% and 70%). Among those who leave the job, many switch to policy-making positions in the UK government or in international organizations.

The program is part of the movement towards the professionalization of the public sector. As such, the organization's key challenges are related to the attraction, retention and diversity of talent.[9] Attracting gender diversity without sacrificing talent or retention was also the goal of the experiment

[9] The business case for diversity lies in the need for social workers to represent the diverse communities where they operate (Cameron, 2001).

we designed together.

Figure 1 illustrates the timeline of the organization's 2017 nationwide recruitment. The experiment happened between September and November, which is the application period. The hiring process consists of different stages (e.g., interviews), which are conducted in a centralized manner either online or at the organization's head office in London. The overall duration of the hiring process from application to job offer is around ten weeks. If a person was hired and accepted the job, actual work in local authorities started in July 2018. I followed workers for the entire duration of the programme until July 2020.

3 Experimental design

Experimental participants are people who are interested in applying for the job offered by the partner organization. To express this interest, potential applicants ("candidates" from here on) should fill in a short registration form on the organization's website which contains eligibility and demographic questions. Completing this form takes between three and five minutes. If eligible to apply, respondents receive an invitation-to-apply email.[10] I introduce exogenous variation in the content of the invitation-to-apply email along two dimensions: perceived gender shares and expected returns to ability.[11] The two experimental conditions were cross-randomized in a fully nested design, leading to a total of four treatment emails. Participants could also be randomly assigned to receive a fifth "pure control" email containing no manipulation, which I used to compare the treatments with business-as-usual for the organization. Randomization was at the individual level, with stratification by gender (man/woman) and ethnicity (white/non-white). The experiment was double-blinded: participants were not aware that the invitation-to-apply email was part of a research study and recruiters were not aware of candidates' treatment assignment. This design limits experimental biases that arise from candidates' knowledge of being in a research study and prevents recruiters' assessment from being influenced by the candidates' treatment. I discuss the experimental manipulations below.

Variation in perceived gender shares. The invitation-to-apply email contained a photograph of a real worker, who was randomized to be either a man or a woman. This experimental condition varies potential applicants' perceived gender shares if seeing a male photograph generates a perception of a higher male share than seeing a female photograph. While this is the main interpretation that I adopt in the paper, photographs may also vary the salience of the predominantly female composition of the job.[12] These two interpretations are observationally equivalent, and I do not distinguish between

[10] The email contains their candidate number, which is necessary to access the application process, and some basic information about the hiring process. Respondents who do not meet the eligibility requirements receive a standard rejection email. Eligible applicants should have a bachelor's degree with a certain minimum average GPA average and have obtained at least a C in Maths and English pre-university qualifications. Social work students are not eligible.

[11] The need to register implies that all the people in the experimental sample are selected on the basis of a minimum level of interest in the job. From a policy perspective, it can be argued that this is exactly the relevant sample to be targeted. Moreover, the brevity of the form and the relatively low application rates after registration (between 50% and 60%) reduce concerns that this sample shows unnaturally high levels of compliance compared to the general population (Al-Ubaydli et al., 2017).

[12] My main interpretation of the photograph's manipulation is aligned with the design of audit studies (Bertrand and Mullainathan, 2004), where non-white sounding names increase the employer's rational expectations that the candidate is going to be non-white. The alternative interpretation based on variation in salience is more aligned with priming studies (Benjamin et al., 2010). This manipulation is also inspired by common business practices.

them, but I will provide manipulation checks that are consistent with the former one.

The reaction of participants to this manipulation identifies the utility given by the workplace gender composition (or related attributes) on their application decisions, assuming that photographs affect choices mainly through changing perceived gender shares. Various confounders might threaten this identification strategy, including ethnicity: if white female candidates apply more after seeing an email portraying a white woman than they do after seeing one with a non-white man, we would not know whether to attribute the effect to the gender or ethnicity match. Moreover, showing photographs of white people right before starting a selection process might cause a stereotype threat in non-white subjects (Steele and Aronson, 1995). For these reasons, I assigned different photographs to white or non-white people and matched the ethnicity of photographed workers with that of each candidate (randomizing gender).[13]

Different elements in the design of this manipulation address candidates' limited attention to the email contents and other potential confounders. To attract the candidate's attention to the photograph, I added a short text where the photographed person addresses the candidate by name and recalls that she/he was also once an applicant. Drawing on studies on role models (Marx and Ko, 2012) and information retrieval (Schwarz et al., 1991), this message should facilitate the candidate's relatability to the portrayed person and the gender group she/he belongs to. The photographed people are real workers who did not feature in other advertising campaigns or multimedia content from the organization. This eliminates unobserved heterogeneity in candidates' exposure to the organization's media channels and recruitment materials. All photographs show the same background and are of the same size to limit visual differences. Other issues might arise if there is a systematic correlation between portrayed workers' characteristics and their gender. I discuss these concerns in Appendix B, where I show that the photographed people are deemed to be similar in characteristics such as friendliness, attractiveness, or work satisfaction by Amazon Mechanical Turk workers.

Variation in expected returns to ability. This type of variation is difficult to induce for several reasons. One would like to communicate to each person what their expected effectiveness on the job will be, given their ability. But ability is imperfectly observed, and this is a new position for the applicants, so no historical data can be used. Moreover, the effect of individualized information on beliefs depends on the level of people's priors, which was unobservable.

I overcome these challenges by providing information about how others performed in the job, allowing participants to infer their returns to ability. To do this, I communicated to subjects the outcome of a selected past cohort of workers, which had either low or high aggregate performance. The exact wording was the following (see Figure 2):

> Did you know that in a past cohort X% of participants got commendable or excellent feedback to their interaction with families?

where X was equal to 66 or 89 in the two experimental treatments. Commendable or excellent are

For instance, Glassdoor advices employers to "Include photographs of women and minority employees on your careers site and Glassdoor. Don't use stock photography" as a strategy for diversity and inclusion (see more here: https://www.glassdoor.com/employers/resources/guide-to-diversity-and-inclusion-in-the-workplace).

[13] To simultaneously test for the effect of workplace gender and racial composition on applications, the ideal design should randomize both gender and ethnicity match/mismatch. However, while not being the main focus of the paper, this would also require a larger sample size.

the highest grades that people can achieve in their performance assessments in the job. In the experiment, these grades referred to the evaluation that workers got when interacting with their customers (i.e. families), so these statistics refer to the social output obtained by previous workers.[14] Both statistics were computed using actual records of the organization. This enabled the communication of truthful but partial information, which on average creates a wedge in beliefs between experimental groups (Dal Bó et al., 2017).

By presenting the job as more challenging (i.e. with 66% rather than 89% of successful workers), a lower past percentage of high performers strengthens the perceived relationship between ability and job outcomes. In contrast, seeing that everyone did well in the past means that there is almost no relationship between ability and outcomes. Lower past success thus signals that talent is rewarded more in the job as compared to a situation in which everyone is successful. Thus, I label the treatment disclosing a low past share of high achievers (66%) as "High Expected Returns to Ability" and the one disclosing an outstanding past performance (89%) as "Low Expected Returns to Ability", which I consider as the default.[15]

This manipulation identifies the effect of expectations of returns to ability under the assumption that performance statistics affect choices mainly through a change in expectations of this parameter. I show manipulation checks in the next section and discuss alternative interpretations in Section 9.2.

I reported information about on-the-job success in frontline interactions with clients for several reasons, primarily to induce variation in people's beliefs concerning their effectiveness in generating output for the employer.[16] Performance metrics on client service are also rarely collected and/or published in the industry, a fact which increases the likelihood that the provided information will affect a candidate's beliefs. Additionally, the quality of clients' interactions is one of the crucial objectives of the organization's mission and it is an important variable that candidates consider when applying (Besley and Ghatak, 2005).[17] Finally, the scores received in practice tasks are the joint outcome of workers' skills and evaluators and clients' reactions. A low score can signal clients' hostility and/or discrimination towards the employees, which can disproportionately affect men and non-white candidates' judgements about their returns on the job (Fisman et al., 2006).[18]

Figure 2 shows an example of treatment email.[19]

[14] Feedback is not given by the families, but by external observers who are experts in social work.

[15] Qualitative interviews conducted with candidates show that 89% was the percentage of high achievers they expected to see, while 66% was surprising to most people. Moreover, application levels in the 89% treatment (52%) was almost the same as in the pure control (53%), further indicating that average priors are closer to 89% than 66%.

[16] Data on job offers were also available, but they could have caused anxiety during the selection process, as shown in studies on information provision before tests (Payne, 1984; Osborne, 2001) and on stereotype threat (Steele and Aronson, 1995), also on white people (Stone et al., 1997). Being long-term outcomes, the chosen statistics can affect beliefs about expected returns on the job, while avoiding negative emotional reactions with implications on short-term performance.

[17] To make this even more salient, the box was positioned below a summary of the organization's mission, which is focused on the challenge of improving outcomes for disadvantaged communities.

[18] Notice that the reactions to the information manipulation would differ if potential minority applicants are trying to infer the likelihood that the employer or customers will discriminate against them. Information indicating low performance could signal that the employer is statistically discriminating, which would generate a negative reaction by men and non-white people. The opposite effect - which is what I find - would thus exclude this interpretation.

[19] From here on, I will denote the four treatment groups by (W,L), (W,H), (M,L) and (M,H), where W or M are for receiving the female or male photograph, respectively, and L refers to low returns to ability (89%) while H refers to high returns to ability information (66%).

3.1 Main manipulation checks

Do photographs and information affect beliefs as planned in the experimental design? I provide manipulation checks conducted on external samples matched on observables with the field participants.

Between November and December 2018, I administered an online survey to 565 people belonging to two samples: 2018/2019 applicants for the same job and workers on the platform "Prolific Academic". The sampling strategy maximizes the similarity to my field sample. Applicants for the same organization in the following year are very similar to my field experiment participants on both observables and, possibly, common unobservable traits that make them interested in this particular job. I selected the sample on Prolific Academic by matching the composition of the field sample on several observables criteria.[20] Both samples were incentivized for participation and the survey had an average completion time of 15 minutes. Appendix B describes the sampling strategy and questions in detail.

In a between-subject design, I randomly assigned respondents to one of the treatment emails used in the field experiment. After mandatory comprehension checks, the survey elicited beliefs on a variety of characteristics of the job, its applicants and workers. Figure 3 shows separately the distribution of answers to the question "Consider 100 people who apply for this job. How many do you think are women?" for respondents assigned to the email with a female or male photograph. The graph shows that the distribution of perceived female shares is shifted to the right in the female as compared to the male photograph treatment. The mean perceived female share is 73.8% and 68% respectively in the two groups (p-val < 0.001). This is consistent with the interpretation of the photograph treatment in terms of a shock in perceived gender shares. In Appendix B I show evidence against confounders related to differences between photographs (e.g., work satisfaction or attractiveness of the portrayed subjects) as well as to other types of information that photographs might convey (e.g., discrimination).

Testing whether people update expected returns to ability in the job requires two ingredients: knowing their approximate position across the ability distribution and the corresponding expected returns. The left panel of Figure 4 shows the distribution of answers to the question "How do you expect a person with your skills and experience to perform in interacting with families in need?" on a scale from 1 (min) to 10 (max). The graph shows that there are no differences in the distribution of answers between the two information treatments, which suggests that people do not change what they think their job-specific ability is as a result of the experimental manipulation. I can then use this question to classify people into low (high) ability depending on whether their answer is below (above) the median.[21] The right panel of Figure 4 shows mean answers to the question "Consider

[20] I selected participants on Prolific Academic to match the share of people in full time employment, who studied subjects related to social jobs and who are of non-white ethnicity in my field sample. All people are from the UK and aged between 18 and 64.

[21] The downside of classifying people's ability based on self-reported measures is that they might strategically inflate their scores (from demand bias, if real applicants think that the employer will see their answers) or be overconfident. These issues, however, become problematic only to the extent that individual misreporting or overestimation alters the ranking of abilities in the sample. The literature on overconfidence reports mixed results on this possibility (see Moore and Healy, 2008; Coffman et al., 2019). Moreover, the manipulation checks reported in this section are still valid even in the case of altered ranking across people as long as the self-reported ability is an accurate measure of the beliefs that drive people's choices. Finally, results are qualitatively similar when using a more objective measure of ability, which is correlated with performance on the job: receiving a first grade in university.

100 people who are applying for this job. Based on the ad you just viewed, on a scale from 1 (worst) to 100 (best), how would you rank yourself for the job among them?", by information treatment and ability level. There are two main takeaways from the bar chart. First, the difference in mean ranking between the 66% and 89% information treatment is negative for low ability applicants, indicating that they expect to be less successful when there are fewer high achievers in the job (difference = -5.70, one-sided p-val=0.03). Secondly, the difference in mean ranking between two treatments is positive for high ability applicants, indicating that they expect to be more successful when there are fewer high achievers in the job (difference = 3.01, one-sided p-val=0.11). Overall, these differences imply that respondents perceive the job to have higher returns to ability when reading the statistic that 66% of people in the past were high achievers than they do with the 89% statistic, as demonstrated by the larger difference in expected rankings between high and low ability people in the former case.

The separate identification of the effect of perceived gender shares from expectations of returns to ability in a fully nested design requires the interaction between the two treatments to be negligible. Data from the auxiliary surveys provide supporting evidence for this requirement. First, respondents' perceived gender shares are not different in the two information treatments. Secondly, updating on success on the job and expected returns to ability go in the same direction independently of the photograph received (see Appendix Figure B.1). Appendix B rules out alternative interpretations of the information provided, such as updating on job amenities (e.g., wage, promotions, training quality).

4 Theoretical framework

In this section I propose a simple model of individual job application where the employer's messages affect expectations of returns to ability ("expectations effect") and utility from job gender composition ("gender shares effect"). The main goal is to guide the empirical analysis and generate predictions, for different parameter ranges, on the size and quality of the applicants' pool in each treatment.

4.1 Environment, preferences and beliefs

Potential applicants are characterized by group belonging g and ability a_i. Everyone can observe their own and others' group $g \in \{M, W\}$, where M stands for men and W for women. Individual ability level a_i is private information, with $a_i \sim U[.]$.[22] Potential applicants decide between applying for a female-dominated job or taking an outside option. Utility in the outside option is a linear function of wage and returns to ability, which I allow to differ by gender: $U^o(a_i) = w_g^o + v_g a_i$.[23] Utility on the job is given by a taste component, which is a function of job gender composition, and expected monetary and non-monetary returns, which are a function of wage and ability:

$$U^j(a_i) = \alpha_i s_g + w + \theta_g(a_i - \hat{a}_g)$$

[22] The assumption that ability a_i is known can be relaxed and replaced with an unbiased expectation of ability. Different transformations of ability are also possible (e.g., coming from overconfidence) and do not affect the theoretical predictions as long as they do not alter the ranking of abilities in the sample.

[23] To ease notation, utility in the outside option does not depend on gender shares. What matters in the model is the difference in gender shares between job j and the outside option, thus s_g can be reinterpreted as the difference in gender share in the two jobs.

where s_g the share of workers of gender g in the job, w is the wage, θ_g are returns to ability and $\hat{a}_g$ is a minimum ability requirement (Lazear et al., 2018). I define the difference between the wage in the job and in the outside option (both known) as $\overline{w}_g = w^o_g - w$.[24]

The utility component $\theta_g(a_i - \hat{a}_g)$ formalizes the fact that agents are motivated by doing a better job than required. This can come from warm glow (Andreoni, 1989), the need to feel competent (Dweck and Elliot, 2005) or internalization of the impact that actions have on the employers' output (Besley and Ghatak, 2005). Alternatively, people might care about impact for extrinsic reasons, if performance is tied to career promotions. Qualitative interviews indicate that both social impact and career opportunities are among people's main motivations for applying. In 2016, the partner organization asked applicants about their motivations for applying: 51% mention career opportunities, 37% mention social impact and 31% mention the challenge of improving local communities. In this view, θ_g can be interpreted as the believed marginal product that a person of gender g with ability a_i achieves in the job and which determines either monetary or non-monetary gains. The parameter $\hat{a}_g$ is the level of ability which is not affected by changes in marginal returns to ability.

The component $\alpha_i s_g$ formalizes agents' utility from workplace gender composition, which is linear in the share of their own gender g. I assume that $\alpha_i \in [0, 1]$, meaning that people prefer working with their own gender and are heterogeneous in this preference. I interpret this preference as a reduced form utility component that can arise from different channels. In my context, threats to identity (Akerlof and Kranton, 2000, 2005), anticipated harassment (Folke and Rickne, 2020) or image concerns (Bursztyn and Jensen, 2017) are especially important for men.[25]

There are two sources of uncertainty in the model. First, agents are unsure of the exact gender share. Their priors are normally distributed $s_g \sim N(\overline{s}_g, \sigma^2_{s_g})$ with $\overline{s}_W > 0.5$ and $s_M = 1 - s_W$. The second source of uncertainty is in returns to ability. Agents do not perfectly know how much reward they are going to get from being above the minimum ability requirement. Priors are distributed differently for the two genders: $\theta_g \sim N(\bar{\theta}_g, \overline{\sigma}^2_g)$, with $\theta_W \perp \theta_M$. I assume that, on average, men think that they have weakly lower job-specific returns to ability in the female-job than women, but they are less certain about this than women.

Assumption 1. ***Gender differences in beliefs about returns to ability***
On average, men believe their returns to ability are lower in the female-job than women: $\bar{\theta}_M \leq \bar{\theta}_W$.

Assumption 2. ***Gender differences in uncertainty***
Men's priors on the returns to ability of both genders are noisier than women's: $\overline{\sigma}^2_M \geq \overline{\sigma}^2_W$.

[24] The organization cannot offer differentiated wages because of the regulation in the sector. I assume that experimental participants know the wage and that this is independent of performance. This assumption comes from the transparency policy of the organization, which publishes the stipend level on the website and a variety of advertising materials.

[25] It is beyond the scope of this paper to micro-found the origin of this preference parameter. Papers in evolutionary psychology (Brewer and Hewstone, 2004) and neuroscience (Eisenberger et al., 2003) show that people fear being in the minority and even feel physical pain when excluded by a group. A rich literature shows evidence of people's preferences for homophily in social networks, including gender similarity. The work by Akerlof and Kranton (2000, 2005) assumes that choosing an activity which is uncommon for one's own group determines a direct loss of utility, either from anticonformism, social exclusion or the cognitive cost of self-image updating (Tajfel and Turner, 1986). When it represents internalized social stigma, the individual component $\alpha_{iM} s_g$ can be micro-founded through a game between applicant i and his peers. In such a setting, α_{iM} is the cost of social punishment for selecting a female job and s_g is the likelihood that the punishment will be enforced.

The combination of assumptions 1 and 2 is equivalent to assuming risk aversion in the utility function and keeping only the assumption of asymmetric uncertainty.[26] Appendix E.2 provides empirical evidence that men tend to have lower and more dispersed expectations of their own group's performance in social work than women. This setting predicts a lower number of men's applications than women at baseline and it builds on a standard Roy model (1951) with perfect correlation between skills in the female-job and in the outside option.

4.2 Reaction to employers' messages: gender shares and expectations

The employer posts recruitment messages to potential applicants in order to increase the gender diversity of applicants.[27] Recruitment messages are vectors containing a photograph P and some information on past performance S, with $p \in \{M, W\}$ and signal $S \sim N(\theta, \sigma_s^2)$, where $\frac{1}{\sigma_s^2}$ is the signal precision and θ is the average return to ability for workers in the job. From here on, I will denote the experimental realizations of the signal $s \in \{s_L, s_H\}$.[28]

The timing of the model is as follows. At time 0, potential applicants know a_i and $\hat{a}_g$ and hold common priors $\bar{s}_g$ and $\overline{\theta}_g$. At time 1, the employer posts ad (P, S). A certain realization (p, s) impacts the individual decision through changes in s_g and θ_g. At time 2, potential applicants decide whether to apply or not given their posteriors on s_g and θ_g.[29]

Pictures $p \in \{M, W\}$ contained in the posted advertisement have a direct utility effect by changing perceived gender shares. Seeing a photograph of gender g increases the perceived share of that gender in the job: $E[s_g|p = g] > E[s_g|p \neq g]$. An observationally-equivalent way in which photographs can affect utility is by changing the salience of gender through α_i, but I do not disentangle these two explanations.

Under the assumption that photographs p have no effect on the way people interpret information, agents form a posterior belief on own returns to ability in a standard Bayesian fashion. Given normality, the posterior θ_g' is a weighted average of the prior and signal s:

$$\theta_g' = \frac{\sigma_s^2}{\sigma_s^2 + \overline{\sigma}_g^2} \cdot \bar{\theta}_g + \frac{\overline{\sigma}_g^2}{\sigma_s^2 + \overline{\sigma}_g^2} \cdot s$$

One of the caveats of predicting people's updating is that both direction and magnitude depend on priors, which are unknown to the researcher. A convenient feature of the experimental design is that identification does not rely on assumptions about priors. As long as the two signals have the same precision and people are Bayesians, random assignment should guarantee that average posteriors on θ_g in the group who received s_H should be higher than in the group who received s_L *independently of priors*. This relies on the following expression for the difference in posteriors between the two

[26] For instance, results go through assuming a CARA utility function in combination with the normality of priors.

[27] Workers' diverse composition might positively affect output through different channels, for instance through skills complementarities (Lazear, 1998), better matching between clients and employees (Hoogendoorn and Van Praag, 2012) or organizational reputation (Miller and del Carmen Triana, 2009).

[28] To design the experiment, I considered the overall mean performance across years as the empirical cut-off for θ and then chose two realizations of yearly performance s_L and s_H respectively below and above the overall mean.

[29] I assume that messages do not affect the knowledge of individual ability a_i (Ashraf et al., 2020; Abebe et al., 2017). This is consistent with the manipulation checks and the fact that information was about aggregate performance and not about people similar to the recipient.

information treatments:

$$\Delta\theta_g = (\theta_g|s_H) - (\theta_g|s_L) = \frac{\overline{\sigma}_g^2}{\overline{\sigma}_g^2 + \sigma_s^2} \cdot (s_H - s_L) \tag{1}$$

$\Delta\theta_g$ is decreasing in priors' precision and independent of priors levels. This is the identification strategy I will use in the empirical section. Assumption 2 of asymmetric uncertainty by gender implies that men will update more than women when receiving the same signal: $\Delta\theta_M > \Delta\theta_W$.

4.3 Predictions

Potential applicants apply for the female-job if $U^j(a_i) - c > U^o(a_i)$, where c is a small application cost. Application choices are fully characterized by ability level a_i. Under a single crossing condition, the decision rule defines a unique threshold of ability a_g^* such that $U^j(a_g^*) = U^o(a_g^*)$.[30] I denote as a_g^* the ability of the marginal applicant. Define $\overline{a}_g$ as the average ability of the applicants' pool of gender g and N_g as its size. Sorting on ability depends on the slope of utilities with respect to ability in the job and in the outside option, which are given by $U^{j\prime}(a_i) = \overline{\theta}_g$ and $U^{o\prime}(a_i) = v_g$, respectively. Lemma 1 states that the marginal applicant is more skilled than the average one when returns to ability in the job are lower than in the outside option.

Lemma 1. ***Relationship between marginal and average quality***
Under the conditions for existence of a_g^*: if $U^{j\prime}(a_i) < U^{o\prime}(a_i)$, then $a_g^* > \overline{a}_g$.

Result 1 states that more applications when people of gender g receive a same-gender ($p = g$) than other-gender ($p \neq g$) photograph identify the effect of gender shares on utility. The quality of such larger pool of applicants is higher when returns to ability in the female-job are lower than in the outside option (negative sorting) and lower in the opposite case (positive sorting).

Result 1. ***The effect of a shock to perceived gender shares***
When $p = g$, the pool of applicants N_g is larger than when $p \neq g$.
If $U^{j\prime}(a_i) < U^{o\prime}(a_i)$: when $p = g$, marginal ability a_g^* and average ability $\overline{a}_g$ are greater than when $p \neq g$.

Let $ds_g = E[s_g|p = g] - E[s_g|p \neq g]$ be the difference in perceived gender shares between receiving a gender matched ($p = g$) or mismatched ($p \neq g$) photograph. The difference in the size of the applicants' pool between the two photographs' treatments is increasing in ds_g, α_i and decreasing in v_g. The top panel of Figure 5 shows the graphical intuition for Result 1. The solid thick line shows the expected utility in the outside option and the two solid thin lines the expected utility on the job, conditional on a certain photograph ($p = g$ or $p \neq g$).

The second result focuses on the effect of a change in expected returns to ability $\overline{\theta}_g$. The effect of this treatment on the size and quality of the pool of applicants depends on two margins. First, whether the marginal applicant has ability above or below $\hat{a}_g$. Second, whether expected returns to

[30] See Appendix E for the formal proof.

ability when receiving a high (s_H) or low (s_L) signals are greater or lower than the returns to ability in the outside option.[31] Define $B = \alpha_i s_g - \overline{w}_g - c - v_g \hat{a}$.

Result 2. ***The effect of a shock to expected returns to ability***
If $B > 0$: when $s = s_H$, the pool of applicants N_g is larger than when $s = s_L$.
If $U^{j'}(a_i) < U^{o'}(a_i)$ and $B > 0$: when $s = s_H$, the pool of applicants N_g is larger and marginal ability a_g^* and average ability $\overline{a}_g$ are greater than when $s = s_L$.

When priors on the returns to ability in the female-job are lower than returns in the outside option, Result 2 shows that raising expected returns to ability improves the average quality of the pool of applicants. The bottom panel of Figure 5 shows the graphical intuition for Result 2.[32] Condition $B > 0$ limits the result to the case in which the marginal applicant has ability level above the minimum ability requirement $\hat{a}$ when $s = s_L$, thus an increase in returns to ability increases its job utility.[33]

The difference in utility between the treatment providing $s = s_H$ and $s = s_L$ is proportional to the change in beliefs between the two conditions $\Delta\theta_g$. A straightforward implication of Bayesian updating is that people with the weakest priors will update the most when receiving new information. This comes from the fact that $\frac{\sigma_s^2}{\sigma_s^2 + \bar{\sigma}_g^2}$ is decreasing in $\overline{\sigma}_g^2$. The implication is that, ceteris paribus, updating will be stronger for men than women because of their higher $\overline{\sigma}_g^2$.

In sum, an increase in the perceived share of own gender in the job can increase applications, but the ability level of the pool of applicants depends on the nature of sorting in the job. Changing expected returns to ability can potentially improve the quality of applicants when there is either positive or negative sorting in female-jobs. The following paragraph summarizes the main empirical predictions.

Empirical predictions

1. Application rates by men and women are larger when candidates receive a photograph of the same vis-á-vis a different gender. Differences in quality between the two treatments are ambiguous.

2. Men's application rates are larger when candidates receive information on higher vis-á-vis a lower expected returns to ability. Differences in quality between the two treatments are ambiguous.

3. The difference in application rates with information on higher vis-á-vis a lower expected returns to ability is greater for candidates with more uncertain priors or with higher job-specific ability.

[31] Notice that the posterior expected returns to ability when receiving s_H could be higher than v_g and the posterior expected returns to ability when receiving s_L could be lower than v_g. I only consider the case in which posteriors when receiving either signal are both higher or both lower than v_g. This means that the change in returns to ability is small enough not to invert the sign of the difference $\overline{\theta}_g - v_g$.

[32] The Theoretical Appendix shows the alternative case $U^{j'}(a_i) > U^{o'}(a_i)$.

[33] If there is negative sorting ($U^{j'}(a_i) < U^{o'}(a_i)$) and $B < 0$, then an increase in returns to ability θ_g discourages the marginal candidate, whose utility decreases because of the increased job difficulty. Also notice that the only source of variation in the sign of B is the level of $\hat{a}$. If $\hat{a} = 0$, the conditions for the existence of a_g^* imply that B is negative if $\theta_g > v_g$ and positive if $\theta_g < v_g$. This means that the quantity and quality predictions of Result 2 do not depend on B if returns to ability are positive for everyone (for $\hat{a} = 0$).

5 Sample, balance and empirical strategy

The experimental sample consists of 5417 candidates, of whom 1013 are men. Table 1 presents summary statistics by gender and balance checks for the overall experimental sample. Candidates' average age is 27 and 3 out of 10 are ethnically non-white. Approximately 32% of the candidates studied in a top-tier UK university, but at the same time the share of people from lower socio-economic backgrounds is substantial, with 19% of subjects coming from families where parents have an unskilled occupation, 27% of subjects receiving economic support in school and 2% being looked after by a social worker as a child.[34] Almost half of the sample (41%) currently work full time (FTE from here on), mostly in the public sector or healthcare, but also in science, business or technology.

Men and women tend to have a similar socio-economic background and experience with the organization, but differ in terms of demographics, education and employment. Men tend to be older and, therefore, more likely to have graduated before 2016 or to be in FTE. The same proportion of men and women attended a top UK university or got a first grade, but men are more likely to have studied scientific subjects and, if working, to be in corporate, scientific or business jobs. The average application rate across years is 60% of registered candidates and it is higher for women than men (by 5 to 10 pp).

Table 1 also shows that treatment assignment is balanced on observables. Columns 7 and 8 report the F-statistics and the related p-value of a regression for each of the row-variables on the set of four treatment indicators. The last column of Table 1 reports the minimum p-value of pairwise t-tests for the difference in means between each pair of treatments along the 23 variables reported.[35] I also fail to reject the null hypothesis of zero effect of all the variables reported in Table 1 in a joint test of orthogonality on assignment to any treatment group (F(23, 4865)=0.67).

Comparing the experimental sample with a random subsample from the UK Labour Force Survey (LFS) with the same age distribution (see the Online Appendix, Table OA.1), I see that men and women in my experiment are selected on interest in public sector or healthcare jobs, a fact which has implications for the interpretation of the empirical results. First, it might indicate that the sample is selected on the weight given to gender shares α_i or priors on $\overline{\theta}_g$, which are the parameters targeted by the experiment. Men in the sample might care less about gender composition or know more about expected returns to ability in social work than the average male LFS respondent (as suggested by the likelihood of being employed in healthcare). This should bias downward my estimates of the effect of varying perceived gender shares as well as of providing information. Secondly, participants in the experiment might have different outside options than average LFS respondents (differing in parameters such as v_g or w_g^o). This implies that selection on talent could be different in other samples facing different structural parameters. Nevertheless, this is a relevant sample for policy and, conditional on interest in the sector, the experimental pool is representative of job applicants to similar programs.[36]

[34] I define top-tier universities as those belonging to the Russell group (see here for the profile: https://russellgroup.ac.uk). In 2012, the total number of people being looked after by a social workers represented 0.12% of the total UK population between 16 and 25 (see the "2012 Care leavers in England data pack" by the Department for Education).

[35] For the few variables with a significant minimum p-value, only one difference out of ten is significant, with the exception of "Young carer" (for which 3/10 comparisons are significant).

[36] For instance, people in my sample resemble applicants for Teach For America (Coffman et al., 2017).

5.1 Main specifications and identification assumptions

In the following sections, I compare the impact of one type of photograph (or piece of information) against the other in encouraging men to enter social work and the impact on the type of applicants attracted into the job. Then, I turn my attention to how new hires originally attracted by different recruitment messages perform on the job and on their retention. My empirical strategy relies on the independent random assignment of the photograph and information manipulations.[37] I estimate the individual-level equation:

$$y_i = c + \beta_1 Pic_i^M + \beta_2 Returns_i^H + X_i'\lambda + \epsilon_i \tag{2}$$

where Pic_i^M is equal to one if i was assigned to receive a male photograph (zero for female photograph) and $Returns_i^H$ is an indicator variable for the high returns to ability information (zero for low returns to ability). The vector of controls X_i contains a dummy for non-white ethnicity (stratification variable), whether the person applied in the past and whether the person registered before the official opening date. As randomization was at the individual level, I use Eicker-Huber-White robust standard errors.

The coefficient β_1 tests the null hypothesis of no effect of perceived gender shares on outcome y_i and measures how this outcome changes in the male photograph vis-á-vis the female photograph treatment. The coefficient β_2 measures how outcome y_i changes in the high expected returns to ability vis-á-vis the low expected returns to ability treatment. Coefficients β_1 and β_2 identify the causal effect of gender shares and expectations, respectively, under the assumption of no interaction between the two manipulations.[38]

I consider two primary sets of variables y_i: applicants' outcomes (entry and quality) and workers' outcomes (performance and retention). *Applicants' outcomes*: While candidates have to pass through different stages of selection, empirically I am only interested in the cumulative effect of the treatment on the candidates' decision to enter the job over the entire process. Accordingly, my outcome variable for application takes the value of one if a candidate decides to submit the application form and shows up at any later stage of the selection process, conditional on being admitted to that stage.[39] To assess whether the treatment attracts better-skilled or worse-skilled applicants, I consider whether i receives a job offer (conditional on application) and whether s/he accepts the offer.
Workers' outcomes: new hires are continuously evaluated throughout the duration of the two-year programme. I measure performance in the job for worker i as the weighted average test score, where weights are given by the credits that the organization assigns to each test. Performing well on these tests is important to be able to qualify as a social worker as well as for future career opportunities.[40]

[37] The comparison with the pure control email is hard to interpret because each treatment email simultaneously changes information and photographs. For instance, the simple addition of creative contents to email advertising can modify consumers' behaviour (Gonzalez and Loureiro, 2014; Bertrand et al., 2010). I thus only compare treatment emails with each other, leaving aside the pure control email.

[38] I show that this assumption is empirically valid for men in Figure A.6, but this exercise has to be taken with a grain of salt because the study is underpowered to look at the interaction.

[39] As shown in Figure 1, the first stage consists in the submission of an application form, which requires motivational questions, list of qualifications and employment experience and an online test on verbal reasoning. The average application rate across years is 60% of registered candidates and it is higher for women than men (by 10 pp).

[40] Assessments include both theory assignments (e.g., case studies, essays) and practice evaluations. The theory as-

To measure retention, I define an indicator for quitting the programme early.

To be able to interpret differences in these outcomes as the causal effect of the treatment on the composition of the pool of applicants, the identification assumption is that the individual probability of being successful from one stage to the next is independent of treatment assignment. This assumption could be violated if treatment assignment influences the employer's screening criteria or candidates' effort. While the double-blind design mitigates these concerns, I also provide evidence consistent with this identification assumption in Section 9.1.

To check for the robustness of the results, I use randomization inference. This method has been increasingly recommended to analyse data from randomized experiments, especially in small samples (Young, 2018). The main idea is that there is some chance that a treatment-control difference would arise because of the units assigned to the treatment group, even if the treatment has no effect. Randomization inference re-assigns the treatment status at random for many repetitions and computes the probability of differences of various magnitudes under the null hypothesis that the treatment had no effect.

6 Results

In this section I show the main results of the paper. First, contrary to common wisdom, I find that gender shares do not affect men's application. In contrast, raising expected returns to ability encourages more men to apply. Crucially, information on higher returns to ability also attracts more qualified male applicants, who get more job offers, perform consistently better on the job for two years and are not more likely to leave compared to men with lower expected returns to ability.

6.1 Application and selection process

The first three columns of Table 2 summarize men's journey in the application and selection process using my main empirical specification (2). The dependent variables in columns (1), (2) and (3) are indicator variables for applying, receiving a job offer (conditional on applying) and accepting the job offer (conditional on receiving the offer), respectively.[41]

The first result is that the workplace gender composition does not affect men's applications. This is shown in the first row of Column (1) in Table 2, which reports the coefficient on the dummy Pic_i^M. Receiving an email featuring a male worker reduces men's applications by 1.7 percentage points with respect to an email featuring a female worker, but this coefficient is imprecisely estimated and I cannot reject the null hypothesis of no difference between the two photographs (see also Figure 6a).

Nevertheless, men react to the expectations manipulation. This is shown in the bottom row of Column (1) in Table 2. The coefficient on the treatment dummy $Returns_i^H$ shows an increase in applications of 7 percentage points in the treatment with higher expected returns as compared to

sessments are evaluated by experts in the sector in anonymous form. Anonymity is not possible in the first month performance review and the practice assessment. The former is a score given by teachers at the end of the mandatory classroom-based training phase which evaluates the fitness and potential of each worker to do a good job in interacting with families. The practice score is given through direct observation of the way in which a worker interacts with customers. Evaluators were not aware of candidates' treatment.

[41] These variables are defined in Section 5.1.

lower expected returns (p-val = 0.04). This effect represents 14% of the mean application rate in the low expected returns treatment and 13% of the pure control mean. This effect could be caused by men applying more with high expected returns to ability or applying less with low expected returns to ability. I qualitatively distinguish these two hypotheses by comparing application rates in each treatment with the pure control group. As the application rate in the low expected returns to ability treatment is close to the pure control group (52% vs 53%), I can conclude that men's entry is positively affected by higher expected returns to ability and that, on average, people hold pessimistic beliefs on the returns to ability in social work.

Importantly, male applicants in the high expected returns treatment are better than applicants in the low expected returns treatment on a variety of observable characteristics that are positively correlated with receiving a job offer. I construct an index which averages the following variables: having a first grade in university, being from a top tier university, having volunteered frequently in the past, having cognitive skills above the median and having obtained the maximum score in English pre-university qualifications.[42] Figure 7a and Table D.1 show that the distribution of this index in the high expected returns treatment is shifted to the right of the distribution in the low expected returns treatment (KS test p-val < 0.10). The positive gap between the two treatments is positive across the distribution, but slightly higher in middle quantiles.

Perhaps not surprisingly, the better observable skillset of men in the high expected returns to ability treatments grants them a higher job offer rate. Column (2) of Table 2 shows that male applicants in the high expected returns treatment also get more job offers than applicants in the low expected returns treatment. The male offer rate in the former treatment is 16%, which is 6 percentage points higher than in the treatment providing information on low returns to ability.

While one might be worried that talented men would not end up accepting the job offer because they have better outside options, Column (3) of Table 2 reassures that this does not happen in my sample. Male offerees in the high expected returns to ability treatment are equally likely to accept the offer than their peers with lower expectations of returns to ability.

As a robustness check, Appendix Table A.1 shows that the difference in men's applications between the two information treatments is nearly the same when combined with a male or a female photograph. This means that the additivity assumption used in my main empirical specification is appropriate.

To sum up, raising expected returns to ability attracts more and better male applicants, who are consequently also more likely to be hired. I now look at the outcomes of male newcomers in the job.

6.2 On-the-job outcomes

The last three Columns of Table 2 show the effect of the treatments on on-the-job outcomes for new male hires, using my main empirical specification (2). The sample in this set of regressions is the subset of job offerees who accepted the offer (43 out of 67 offerees) and started working for the organization in July 2018 (see Figure 1). After a first month of training, new hires are sent to their allocated team

[42] All the variables are first standardized to be mean zero and unitary standard deviation in the pooled sample of men. To define cognitive skills, I use the employment history reported by each applicant in the application form. I coded the most recent reported role into standardized SOC4 categories and followed the methodology of Acemoglu and Autor (2011) to match each occupation with the skills listed by O*Net.

in different UK regions.[43]

Columns (4) and (5) of Table 2 show that raising expectations of returns to ability allows the employer to select male workers that are not only better in terms of observable qualifications, but also perform consistently better on the job. The average difference in average test scores between the two information treatments is 4.8 percentage points over the two years on the job, which corresponds to 8.6% of the mean average score in the low expected returns treatment. The average performance of men with higher expectations of returns to ability is 9.5% percent higher than the performance of men with lower expectations during the first semester (Column (4) of Table 2). This gap slightly decreases to 8% in the last year of assessments, but on an average level of performance which drops substantially and becomes more heterogeneous.

Male workers attracted to apply by higher expected returns to ability perform better along the entire distribution of on-the-job performance. Figure 7b shows the distribution of men's residualized average test scores by experimental treatment, after controlling for the basic set of controls from specification 2. The distribution of average test scores is completely shifted to the right for men in the high expected returns vis-à-vis the low expected returns to ability treatment (right-hand side figure, KS test p-val < 0.05).[44]

On-the-job performance is good for the employer and the service beneficiaries, but it might come at the expenses of workers' job-satisfaction. Newspaper headlines often report high levels of burnout, stress and, consequently, turnover among social workers (Ravalier, 2018). In the UK, as many as 50% of social workers plan to stay less than two years in the occupation and work overload is one of the most common reasons for leaving (Ravalier, 2018). Thus, one might worry that newly hired high performers would soon become overloaded with cases and/or be more likely to leave.

However, Column (6) of Table 2 shows that men in the high expected returns to ability treatment are not more likely to drop out of the programme. Furthermore, Table 4 shows that male workers with higher expected returns to ability are more likely to state that they want to keep working in social work after the programme and have more positive attitudes towards their work. They are more likely to say that they are aware of the social impact of their work and that they are confident of their skills in interacting with the families. At the same time, Columns (1) and (2) show that men with higher expectations of returns to ability are not more likely to be concerned with the work environment or to have a higher perceived workload.[45]

To show the robustness of the performance results and overcome empirical issues related to the small sample, I exploit the availability of repeated tests for each worker by estimating the model:

[43] Allocation is based on individual regional preferences, slot availability and diversity considerations. The organization tries to satisfy individual preferences in most of the cases: out of the ones who accepted the offer, 70% were allocated to the first ranked region. There are a total of 52 communities in my sample and the average team size is 4 people.

[44] Figure 7b also shows that the distribution of average test scores of men assigned to the male photograph email is slightly shifted to the left vis-à-vis recipients of the female photograph (left-hand side). This evidence is consistent with slightly higher men's applications in the female photograph treatment in a model where men are negatively sorted in the job, but contrasts with the better observable skillset if men in the male photograph treatment. Thus results are more mixed regarding selection on talent in the photograph manipulation.

[45] They also have the same actual workload than men in the lower expected returns to ability treatment. The data used in this table come from surveys conducted by the organization every six months among workers. The data use three rounds of surveys and all the regressions control for wave dummies.

$$score_{ia} = \alpha + \beta_1 Pic_i^M + \beta_2 Returns_i^H + X_i'\lambda + \epsilon_{ia} \quad (3)$$

where $score_{ia}$ is worker's i grade in assessment a normalized by the mean and standard deviation of male workers' grades.[46] The vector X_i includes, in addition to the basic controls of specification (2), dummies for the region where the worker is allocated and a dummy for whether the worker has been allocated to his preferred region. I also estimate the same model controlling for the index of desirable skills introduced in section 6.1 and an index of difficulty of the local community where a worker is allocated to.[47] Standard errors are clustered at the worker level.

This specification compares the score achieved in a given test by two men working in the same (preferred) region - and assigned to a local community of similar difficulty - who were originally assigned to receive different photographs or pieces of information at recruitment. Coefficients β_1 and β_2 measure the causal effect of the experimental manipulations on the selection of more talented workers under the identifying assumption that the treatments do not have a direct impact on workers' effort and/or motivation on the job.

The bottom row of Table 3 confirms that men with higher expected returns to ability perform significantly better: their scores are 0.25 standard deviations higher than men with low expected returns (p-val < 0.10) and the effect slightly increases when controlling for the difficulty level of local communities (Column (3)). Once I control for workers' observable skillset, the coefficients on the high expected returns treatment dummy decrease by two percentage points and become marginally insignificant (F-stat = 1.56). This means that the observable measures of talent used by the employer to make job offers partially account for the difference in performance between the two treatment groups, which is consistent with my interpretation of the treatment affecting job performance through selection. The residual difference can be due to a wider range of skills which are not observable to the researcher (but which the employer is able to pick-up in hiring) or to a direct impact of the treatment on effort on the job. I provide evidence against the latter channel in the following section.

The top row of Table 3 also shows that men attracted through a male photograph perform worse than men attracted through a female photograph. Men in the former treatment show 0.28 standard deviations lower scores than men in the latter treatment (p-val > 0.10), an effect which is enhanced when controlling for ability measures (p=0.10). This result suggests that, despite the null effect of the photographs at application, the male photograph attracts men with better formal qualifications, but who end up being worse matches for the job.

All in all, the evidence of this section is consistent with a story in which higher expected returns to ability improve the male workforce through selection. Higher expected rewards to skills attract a better pool of male applicants, from which the employer selects talented hires who also perform better once in the local communities.

[46] In the raw data, each grade is on a scale between 0 and 100, where 40% is the minimum threshold for passing.

[47] I use data from the Department for Education on the Children and Family Social Work Workforce (2017) in England and data from the 2016/17 report of Her Majesty's Chief Inspector of Education, Children's Services and Skills (by Ofsted). For each local authority, I compute an index of "difficulty" by averaging the score in social workers' caseload, turnover, absenteeism and Ofsted's scores on helping children, childcare, leadership effectiveness.

Discussion

The null effect of the photograph manipulation on men's application is surprising. Adverts portraying people of the same gender are a key ingredient of most policy proposals trying to attract men to teaching or nursing (Flynn, 2006; The Economist, 2018).[48] For instance, in 2002 the Oregon Center for Nursing tried to appeal to young men by launching the notorious "Are you man enough to be a nurse?" recruitment campaign, which portray a line-up of masculine men engaged in a variety of extreme sports.[49] A more realistic representation of male nurses is also one of the pillars of the biggest recruitment drive in the history of the UK National Health System (NHS from here on, see McGonagle, 2019). In light of my results, it's surprising that the campaign has been considered the cause of the 9% increase in men's enrollment in nursing school between 2018 and 2019 (News, 2019).[50]

One way to reconcile my results with current policies is thinking about sample selection. If men in my sample care less about workplace gender shares than the average man (low α_i) , the estimated effect of the male photograph provides a lower bound of what we should expect in the wider male population. However, I run an additional experiment with the same organization to show that adverts portraying men are ineffective in encouraging even a wider population of male students to apply (see the Online Appendix, Section B). This extends the external validity of the null result of the photographs and suggests that the gender composition among campaigns' actors does not matter for men's recruitment. What matters instead is the informational content of the campaigns, as demonstrated by my second result.[51]

Information provision has an economically substantial effect on men's applications. Previous field recruitment experiments which exogenously varied on-the-job incentives find effects of higher wages on application rates to be between 18% and 26% (Abebe et al., 2017; Dal Bó et al., 2013). In my experiment, raising expected returns to ability increases applications by 13% with respect to the pure control group mean application rate. This effect is between two thirds and half of the one obtained in the aforementioned papers, but it's also nearly costless for the employer.[52]

How can information have such a big impact? According to Bayesian updating, the magnitude of the coefficient on the $Returns_i^H$ dummy is proportional to men's uncertainty in priors. Thus the large gap in application rates between the two information arms uncovers an important barrier to men's entry in female-dominated jobs: informational constraints. While it is surprising that limited information plays a role in my context, where one could assume there are nearly unlimited oppor-

[48] See, for instance, the article "Male Nurses: not just a woman's job" in The Economist (August 2018) here: https://www.economist.com/britain/2018/08/18/a-shortage-of-nurses-calls-for-the-recruiting-of-more-men.

[49] See the webpage: http://www.oregoncenterfornursing.org/index.php?mode=postersandmore.

[50] See more at these links: www.england.nhs.uk/2019/02 and www.nhsemployers.org/your-workforce/recruit/employer-led-recruitment/we-are-the-nhs-campaign.

[51] The null effect of gender composition on men's applications is in line with estimates by Hsieh et al. (2019), who find little room for occupation-specific preferences in explaining changes in the allocation of talent in the last decades, and with Wiswall and Zafar (2018), who show that neither men nor women are willing to receive a lower wage to work alongside a greater proportion of people who share their gender.

[52] A similar light-touch intervention is the one by Del Carpio and Guadalupe (2018), who get around 28% higher applications by women in the treatment group. However, this gap can be explained by differences in the application costs between settings (completing the application form takes ten times longer in my setting than in Guadalupe and Del Carpio's paper) as well as the level of application rates in the control group (53% in my context vs 7% in their setting).

tunities for learning and experimentation, the willingness to experiment is itself a function of the expected usefulness of information. The sheer fact that some occupations are almost exclusively done by women can impair men's inclination to collect - or simply pay attention to - information on careers that are uncommon for their gender. This reluctance to get informed might be especially detrimental to people with high job-specific talent and a valuable outside option, who thus could benefit the most from information provision as shown by my results on applicants' and workers' talent.

To the best of my knowledge, this is also the first paper showing that disclosing a lower past success rate can be motivating for new applicants. This contrasts with many role model interventions, whose standard design provides high statistics of success to minority members to increase their perceived likelihood to succeed in jobs which are uncommon for their demographics (see, e.g., Del Carpio and Guadalupe (2018)). The insight that I add to these studies is that a high probability of success might be interpreted as signal of low returns to ability rather than the unconditional probability of success, which might encourage only people of low ability to apply for the job. This is especially likely to happen in sectors where people's baseline priors on returns to ability are low, such as social work, but previous research in this area has only focused on entry in male-dominated occupations where priors can differ substantially.

A final point worth discussing is that my information treatment does not provide information on incentive schemes or earnings, but about the extent to which talent can be impactful and effective in the job. Highlighting the valuable contribution that men can make in female-dominated jobs is a common policy content (Flynn, 2006), but there have been no previous attempts to link these contents to the rich literature on inferences about returns to ability and educational/occupational choices (Nguyen, 2008; Jensen, 2010; Zafar, 2013; Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015, 2018; Boneva et al., 2017).[53]

7 Theory-driven heterogeneous treatment effects

7.1 Heterogeneity by gender norms and priors' uncertainty

Does gender composition or expectation of returns to ability matter relatively more for men not used to seeing other men in the job? The model predicts that the impact of a change in gender composition is increasing in individual taste parameter (α_i) and that the impact of new information is increasing in initial uncertainty on job returns ($\overline{\sigma}_g^2$).

I build an individual-level measure of exposure to labour market gender segregation *during teenage years* as an empirical proxy of the individual weight on gender composition α_i and uncertainty of men's returns in female-jobs $\overline{\sigma}_M^2$. A rich literature shows that segregation is associated with social norms of what are appropriate activities for men and women (Goldin, 2014; Cortes and Pan, 2018; Baranov et al., 2020). Exposure to gender segregation can also affect the persistence of biased beliefs on

[53] Preferences for competitive environments are another way in which returns to ability may enter the individual decision problem (Niederle and Vesterlund, 2007; Wozniak et al., 2009; Dreber et al., 2014; Flory et al., 2015; Reuben et al., 2017). Even if non-competitive types usually sort into female-dominated occupations (Buser et al., 2014), in other contexts we might expect incorrect inferences about returns to ability to be amplified through their interaction with preferences for competition.See also the discussion in Section 9.2.

group ability, an insight used by Arrow (1973, 1998) to explain the persistence of long-term statistical discrimination.[54] I posit that a similar channel can limit minorities' knowledge of their own returns to ability in uncommon jobs.

My proxy for traditional gender norms and uncertainty on men's returns in female-jobs exploits heterogeneity in the geographical origins of candidates. Using microdata from the 2011 U.K. Census, I construct the Duncan index of occupational segregation (Duncan and Duncan, 1955) of the local area where a candidate went to secondary school and/or lived during teenage years.[55]

Panel A of Table 5 estimates heterogeneous treatment effects on applications by splitting the sample between subjects exposed to higher-than-median (Column 1) and lower-than-median (Column 2) occupational gender segregation. The top row shows that exposure to occupational gender segregation does not mediate reaction to photographs. In contrast, the bottom row of Panel A shows that men exposed to higher-than-median occupational gender segregation react significantly more to the high returns to ability information. Their applications increase by 16.5 pp, which represents 34% of the mean in the low expected returns group. This suggests that occupational gender segregation can affect men's choices of occupations through a limited information channel, which increases their uncertainty and/or biases in beliefs about gendered returns to different occupations.[56]

Appendix C contains more details on the methodology and presents additional exercises. First, I designed and implemented an ad-hoc Implicit Association Test (IAT) to show that exposure to segregation increases the automatic association between social work and women. Secondly, using data from the British Attitudes Survey and the World Value Survey, I show that UK regions with high gender segregation levels display more traditional norms related to women's employment. Third, using auxiliary online surveys, I show that men coming from areas of with a high Duncan index tend to have higher uncertainty in beliefs about men and women's abilities in female-jobs.

7.2 Heterogeneity by outside option parameters

The joint increase in male applications and hires is consistent with a model where potential applicants face steeper returns to ability in the outside option than in the job (case $\theta_M < v_M$). In this section, I

[54] In Arrow's words (1998, p.97): "To the extent that discrimination takes the form of segregation, then there will in fact be little experimentation to find out abilities".

[55] The Duncan index identifies the percentage of women (or men) that would have to change occupations for the occupational distribution of the two genders to be equal and is computed using the following formula: $\frac{1}{2}\sum_{i=1}^{N}|\frac{m_i}{M} - \frac{f_i}{F}|$, where m_i and f_i are the male and female population, respectively, in occupation i and M and F are the total working population of the local labour market. It takes values between 0 (complete integration) and 1 (complete segregation). Using a bridge, I merged the index with my experimental data through the subjects' secondary school postcode and, when missing, home postcode. I use the current location for the 62% of people on whom I have no data on the secondary school location. For students (who are 50% of these missing cases), home location is the parents' residence, which is thus a proxy of where they grew up. For workers, it is the current domicile. Results are qualitatively the same running the same set of regressions of Table 5 using only the subset of people with data on school location, but power drops.

[56] The main caveat for the interpretation of Panel A of Table 5 is that there might be omitted factors which vary by exposure to job genderization which confound my estimates, but results are unchanged by the inclusion of controls for observable differences between men coming from areas with high versus low gender segregation. Results are also robust to the inclusion of a regressor for the ratio of male to female unemployment at the local area, to control for possible confounders in terms of gender differences in working opportunities. Columns (3) and (4) of Panel A of Table 5 repeat the same exercise using a different index: the average share of men working in female-dominated occupations in the local labour market.

test this theoretical prediction empirically.

Using data from the Labour Force Survey, I proxy returns to ability in the outside option with the average wage dispersion faced in the UK labour market. For a candidate who studied subject s, wage dispersion is computed as the weighted average of the 75/25 interquartile range of the distribution of hourly wages across industries in the UK labour market, where weights are given by the proportion of graduates of subject s working in each industry.[57]

Panel B of Table 5 shows heterogeneous treatment effects by splitting the sample of candidates by above/below median wage dispersion faced in the UK labour market. As hypothesized, Columns (5) and (6) of Table 5 show that the difference in application rates between the low and high returns treatment is three times greater for men facing wage dispersion above the median than below the median. This confirms that the marginal applicant induced to apply by the information treatment faces a steeper outside option, a fact which also explains the subsequent increase in average quality of the pool of applicants

Panel B of Table 5 shows further heterogeneous treatment effects with respect to the outside option level (w^o). The last three Columns of Table 5 split the candidates' sample by terciles of the individual outside option, defined as their expected hourly-wage in the U.K. labour market conditional on subject studied, gender, race, age, British nationality and marital status.[58] The evidence reported in Columns (7), (8) and (9) of Table 5 show that high expected returns to ability on the job increase application rates by 11 percentage points among men in the first tercile of the male outside option distribution, an effect which almost halves in higher terciles. This result is consistent with the interpretation of the information treatment as a change in the slope of expected utility on the job.

7.3 Heterogeneity by job-specific ability

In this section I test the theoretical prediction that higher expected returns to ability should attract differentially more applications by high ability people. I use a discrete-choice framework to quantify the change in the slope of application probability with respect to job-specific ability generated by expectations manipulation and to benchmark this effect against a change in wage.

Consider the individual decision of whether to apply to the job or not: $Pr(apply = 1) = Pr(U^j(\alpha_i, s_g, \theta_g, a_i, \hat{a}) + \xi_j > U^o(v_g, a_i, \overline{w}_g) + \xi_o)$, where ξ_j and ξ_o are errors with type I generalized extreme value distributions and the cost of application is assumed to be zero.[59] I use Maximum Likelihood to estimate the following logit model:

$$log\frac{Pr(apply)}{(1 - Pr(apply))} = \beta_1\overline{w}_g + \beta_2 OwnGender_i + \beta_3 a_i + \beta_4 Returns_i^H + \beta_5 Returns_i^H * a_i$$

where $OwnGender_i$ is a dummy for a same-gender photograph, $Returns_i^H$ is a dummy for high ex-

[57] This index of wage dispersion is a function of the endogenous choice of university subject made by the candidates. Thus, Panel B of Table 5 could capture heterogeneous treatment effects due to other unobservable differences between candidates who chose the same university subject. As a robustness check, in the Online Appendix (Table OD.1) I repeat the exercise using the wage dispersion of the region where each candidate lives. Mobility across regional labour markets. is limited: in the LFS, only 16% of workers work and reside in different regions.

[58] I used data from the 2017 and 2018 UK Labour Force Survey. The Online Appendix contains a detailed summary of the methodology used (Section A).

[59] Results are similar when including the distance to London as a proxy for the cost of applying.

pected returns to ability, $\overline{w}_g$ is the de-meaned difference between the log-wage in the job and in the outside option and a_i is a de-meaned proxy of job-specific ability. This proxy is the predicted on-the-job performance score based on observables and obtained from the pure control group through a linear truncated regression. The assumption in the construction of this proxy is that the way in which these observables affect job performance is independent of being hired and treatment status.[60]

The coefficient of interest in the model is β_5, which identifies the difference in slopes with respect to job-specific ability between the two information treatments $\Delta\theta_g = \theta_H - \theta_L$. The estimated β_5 for men is 0.02, which implies that the odds of applying increase by 22% for a one unit increase in job-specific ability. As a benchmark, just above mean ability this coefficient is comparable to a substantial 11% increase in the wage in the job (an increase in the hourly wage from 20 to 22.2 GBP).

Consistently, Figure 9 shows that the slope of the predicted probability of applying with respect to imputed on-the-job performance is higher in the treatment with expectations of higher returns to ability vis-á-vis the alternative information treatment. This is consistent with the interpretation of the information treatment as a change in the slope of expected utility with respect to job-specific ability.

8 Net effect of policies to attract men in pink-collar jobs

Is it desirable to attract more men into pink-collar occupations through a policy which raises expected returns to ability? This Section provides an answer to this question in two steps. First, as any policy can rarely target only one particular gender, we need to know whether attracting more men by raising expected returns to ability also affects women. Knowledge of the sign and magnitude of these spillovers on women's selection will allow me to assess the net impact of the policy for the employer. I conclude by discussing implications for the economy overall.

8.1 Spillovers on women's selection

The first result is that women are insensitive to the experimentally-provided information on returns to ability on average. The second row of Table 6 shows that the coefficient on the $Returns_i^H$ dummy on women's applications is -0.015 and statistically insignificant, while Column (1) of Table A.2 confirms that the two genders react differently to the expected returns treatment. The null effect of information provision on women's applications is consistent with the majority holding more precise priors on occupational returns to ability (assumption 2 of the model).[61] This evidence could lead to the conclusion that raising expected returns to ability is a silver bullet for the employer: it achieves higher diversity, better quality among the gender minority and has no effect on the majority.

[60] The truncated regression uses the following variables: ranking and average completion rate of the university attended by the candidate, subject studied, obtaining a first grade, whether the grade is expected or obtained, age, age squared and whether the person is in FTE. Data on about the university attended by the candidate are taken from the 2015-2016 University and Subject League Tables, which systematically collect public data from the Higher Education Statistics Agency (HESA) and the National Student Survey (NSS). For more information see the webpages: https://www.thecompleteuniversityguide.co.uk, https://www.hesa.ac.uk and https://www.thestudentsurvey.com.

[61] Using the discrete-choice model introduced in section 7.3, the right-hand side panel of Figure 9 shows that the distribution of the estimated change in expected returns to ability $\Delta\theta$ for men is shifted to the right of women's, with a mean $\Delta\theta$ of 0.01 for women and of 0.02 for men.

Nevertheless, we can go a step further. While the suggested policy to attract more men does not affect women's selection, the very fact that men start entering social work might have an impact on women's behavior.[62] If male shares in female-dominated jobs increased, what would be the impact on the number and quality of female applications and hires? I can provide an answer to this question using my photograph manipulation. Showing a male photograph allows me to simulate - in people's minds - a counterfactual world in which the share of men in the job is higher. Thus, women's behavior when receiving a male photograph tells us how women would react if more men became social workers.

I find that increasing social workers' male shares discourages women from applying for the job. Column (1) of Table 6 shows that there are 7.5% fewer women's applications in the male vis-à-vis female photograph treatment.[63] Crucially, the drop in women's applications does not come at the expenses of a lower applicants' quality. On average, the skills of female applicants in the male photograph treatment are not significantly different than in the female photograph arm, and indeed the offer rates and job performance in the two arms are similar (Columns (2) to (5), Table 6).[64]

However, the male photograph affects women's decisions in two other stages of the process: offer acceptance and job retention. Women who applied despite seeing the male photograph are more likely to accept the job offer vis-á-vis women who received the female photograph (Column (3) of Table 6), especially those with lower observable ability (Figure A.5b). However, this greater motivation fades when women find it hard to keep-up with the requirements of the job. Colum (6) of Table 6 shows that women in the male photograph are more likely to leave the programme before completing it. This effect comes entirely from women with low performance in the first six months in the job (Figure 8), of whom 42% belong to the group of women with below-median observable skills. To sum up, these results indicate women with high job-specific talent are relatively unaffected by changing perceived male shares in the job. A higher male share instead affects the decisions of lower ability women, who seem to become more sensitive to either positive or negative signals about their fit with the role either at recruitment or in the job (Coffman et al., 2019).

8.2 Net impact for the employer and the economy

The evidence discussed in the previous section allows me to answer the question of whether attracting more men by raising expected returns to ability is desirable for the employer.

First, if we only look at recruitment, the answer is positive. Information on higher impact of ability attracts more and better men, but does not affect women's applications on average. In turn, a higher

[62] History provides a few examples of female-dominated occupations where the gender composition tipped in favor of men, such as coding and hospital administration (Arndt and Bigelow, 2005; Ensmenger, 2012).

[63] Figure 6b shows application rates across treatment groups from the raw data. An alternative interpretation of this effect is that women think that the employer is looking for men, but this would imply also a positive reaction by men to the male photograph. Moreover, this alternative story does not rule out that the effect is driven by an anticipated future change in gender composition and still shows that an employer's active policies to attract more men in female-dominated jobs might discourage women from applying.

[64] There are distributional differences. The distribution of female applicants' observable skills has a greater variance in the male photograph than female photograph treament (Figure A.5a). At the same time the average applicant's ability in the middle percentiles goes up (Appendix Table D.1), indicating that the relatively least talented women belonging to the middle ranges decide not to apply. In the middle ranges of ability, the joint decrease in applications and increase in quality suggests that the sorting of women in the job is positive. Table A.3 confirms this conjecture by showing that the negative effect of the male photograph is concentrated among women with flatter outside options.

perceived male share does not affect the quality of new female hires and increases their acceptance.

Nevertheless, the desirability of a policy which attracts more men by increasing expected returns to ability is ambiguous when considering the outcomes of the incoming pool of workers. The reason is that a higher male share and higher expected returns to ability discourage low-performing women from staying in the job, creating a trade-off between performance and retention for the employer. Is the increase in performance high enough to justify the loss of female workers in the programme?

I conducted a survey in the organization to provide an answer to this question. I asked the recruitment personnel to choose between two hypothetical cohorts of workers: one with no drop-out, but lower average performance and one with greater drop-out, but higher performance. [65] 58% of respondents prefer a cohort with higher performance rather than longer retention, and this percentage increases to 71% if higher performance is also associated with greater gender diversity. All in all, this evidence indicates that the net effect of the increased-returns-to-ability policy is positive, and that the gains it achieves in terms of diversity and performance justify the costs in terms of turnover. Table 7 provides evidence of the improvements in diversity and performance under different recruitment messages, net of turnover. Providing information on high expected returns to ability, across photographs, achieves the highest overall performance and male share in the workforce.

What do my results imply for talent allocation in the aggregate economy? In a world with two sectors (e.g., social and private), the answer ultimately depends on the nature of men and women's sorting in each. If men's sorting in female-dominated jobs is negative, as my results indicate, their reallocation will improve average skills in both sectors of the economy as switchers have the lowest private-sector ability. Things are more complicated if we consider the effects on the crowd-out of women. There will be aggregate gains from talent reallocation if women are positively sorted in female-jobs and negatively sorted in the outside option, because switchers from female-dominated jobs to the alternative will improve average quality in both. If, instead, women are positively sorted in both sectors the net effect of both women and men's relocation will be ambiguous.

In the longer term, implications for the aggregate economy are difficult to predict because they depend on whether men's entry will generate further effects, such as tipping-point reallocation across occupations (Pan, 2015) or backlash from women (Rudman and Fairchild, 2004).

In the US, Hsieh et al. (2019) show that the increase of women's and black men's shares in high-skilled occupations since 1960 is related to a weakening of group-specific occupational barriers. In turn, this has positive effects on aggregate growth outcomes as the newcomers in high-skilled professions have also high occupation-specific talent. While I do not have data to provide evidence on aggregate effects, my experiment complements this work by showing that men might similarly be facing occupation-specific barriers in female-dominated jobs. Under some assumptions on the correlation of skills in the economy, this implies that men's current allocation in female-dominated jobs is suboptimal and talented male social-workers are not reaping the highest returns to their ability.

[65] To construct the two scenarios, I used data about female workers' outcomes in treatment (W, L) and in treatment (M, H) respectively. Details of the survey are reported in Appendix G.

8.3 Do women care about the workplace gender composition?

Women's reaction to the photograph manipulation could suggest that women value co-workers gender more than men do (Haile, 2012; Lordan and Pischke, 2016). Yet, I cannot reject the null hypothesis that men and women react in the same way to the photograph manipulation (Table 6). Moreover, women are less likely to apply (and to stay in the job) when seeing a male photograph *in combination* with information on high returns to ability. When expected returns to ability are low, the impact of the male photograph on women's decisions is reduced (see Figure A.6). This interaction suggests that women's behaviour does not stem from a generalized preference for working with their own gender.

There are two main ways to interpret women's behavior. A first hypothesis goes through preferences: women dislike working with men in more challenging environments (Niederle and Vesterlund, 2007; Niederle and Yestrumskas, 2008). Nevertheless, this cannot explain the observed differences in behavior by low-performing vs high-performing women, unless preferences are correlated with ability.

A second hypothesis is that gender shares affect women's inference of their expected success on the job, as suggested by previous work (Croson and Gneezy, 2009; Wozniak et al., 2009; Dreber et al., 2014; Coffman et al., 2019; Bordalo et al., 2019). A lower female share might signal a decrease in women's stereotypical advantage in social work. This is bad news for all women, but especially for those of lower ability in a workplace where ability matters more (as indicated by higher expected returns to ability). Empirical evidence supports this interpretation: women's early exit from the job is concentrated among women not only working in teams with a high male share, but who also perform worse than their team peers (see Figure A.3).

Section E.4 adds beliefs based on stereotypes to the model. I assume that gender shares impact expected on-the-job ability by providing information on the minimum ability requirement $\hat{a}$, such that a higher male share increases the perceived minimum ability level needed for a woman. This implies that women who would have applied because of their stereotypical advantage are now required to assess whether their ability level is going to be high enough to succeed. If the effect of gender shares on $\hat{a}$ is strong enough, this model predicts a negative difference-in-difference in application rates between the male and female photograph and in the high versus low expected returns to ability treatments, which is what we observe in the data.

9 Robustness and alternative mechanisms

9.1 Potential threats to identification

9.1.1 Does the treatment affect employer's screening?

To attribute the increased offer rate to the causal effect of the treatment on applicants' composition one needs to exclude the possibility that the treatment affects the employer's screening criteria (Ashraf et al., 2020). I test this assumption using the following specification:

$$o_i = \sum_j \alpha_j^{T^1} T_i^1 X_i^j + \sum_j \alpha_j^{T^2} T_i^2 X_i^j + S_i'\lambda + \epsilon_i$$

where o_i is equal to one if i received a job offer (conditional on applying), T_i^1 and T_i^2 are indicator variables for one of the two treatments for each condition (e.g., male and female photograph respectively) and S_i are the two stratification variables (gender and ethnicity). X_i^j are indicator variables equal to one if candidate i has a certain qualification, such as receiving a first grade, having studied a subject aligned with the content of the job, having received the maximum score in Maths or English pre-university qualifications.[66]

Columns (1) and (3) of Table A.4 report the coefficients $\alpha_j^{T^1}$ and $\alpha_j^{T^2}$ for the information and photograph conditions, respectively. Columns (2) and (4) report the p-value of tests of equality of coefficients $\alpha_j^{T^1} = \alpha_j^{T^2}$. There are two takeaways from the table. First, the employer finds some qualifications more desirable than others. For instance, candidates who received a first grade in university are 11 percentage points more likely to receive an offer, while receiving a high score in Maths does not matter. Secondly, I cannot reject the null hypothesis of equality of the employer's selection criteria across treatments, as coefficients $\alpha_j^{T^1}$ and $\alpha_j^{T^2}$ are statistically indistinguishable in most the cases.[67]

9.1.2 Does the treatment affect candidates' effort?

The higher quality of male job offerees and workers is consistent with better selection generated by the high returns expectations treatment. An alternative explanation of this effect is a self-fulfilling prophecy: believing in higher chances to be successful might encourage men to put more effort into the hiring process or in the job, with a subsequent higher offer rate and performance at work. Such an effect has been documented as a response to varied leaders' expectations (Rosenthal, 1994; Eden, 1992) or to prejudice against minorities (Benyishay et al., 2016; Glover et al., 2017). There are three main pieces of evidence against this explanation.

First, any motivating effect of the treatment should be stronger right after receiving the invitation-to-apply email. In contrast, Table A.5 shows that men in the two information treatments do not differ in the effort put into application completion, as measured by the percentage of fields filled-in and the number of characters used to answer the application questions. Second, male workers in the high expected returns to ability treatment have on average the same actual and perceived workload in the job, which also excludes differential effort in the workplace (Columns (1) and (2) of Table 2). Third, we should expect higher effort to be correlated with higher likelihood of job acceptance, perhaps through a sunk cost fallacy. Evidence reported in Table 2 contradicts this hypothesis.

A related concern is that the performance effects are an artefact of the experimental manipulation and come from a "surprise" once people compare expected and actual returns on the job. There are

[66] I define "aligned subjects" as those with knowledge in the key areas listed by the O*Net website for social work. For instance, O*Net lists "Law and Government" as one of the knowledge components required in the job. I thus classify subject titles containing "Law" and "Government" as an aligned subjects. I also ran the same specification adding measures of cognitive and manual skills and results are robust to this inclusion. The employer selects people with higher cognitive skills, but manual skills are deemed less important and there are no differences in the relevance of these two categories across treatments.

[67] Only two comparisons out of sixteen are significant. The employer is more likely to give an offer to people with a first grade in the male photograph than the female photograph treatment and more likely to give an offer to people who studied a subject aligned with the job in the high expected returns than low expected returns treatment. Importantly, the latter difference is driven by female candidates and thus cannot explain the increase in offer rates seen in the high expected returns treatment for men.

two implications of this hypothesis: performance effects should be waning over time and be driven by people surrounded by worse colleagues. Figure A.2 shows that there is no decreasing trend in the coefficients on the treatment indicator variable in separate regressions for each of the nine on-the-job assessments. I also do not find evidence of a greater performance by men in teams with a lower leave-out-mean in the high versus low expected returns treatment.[68]

9.2 Alternative mechanisms

9.2.1 Social comparison and employer's selectivity

One way in which participants in my experiment could interpret the information provided is by forming expectations about others who are competing for the same role or about the selectivity of the employer. In fact, evidence from auxiliary online experiments (Appendix B) show that respondents think that the proportion of applicants with the potential to be high-achievers in the job is lower when they received the 66% than 89% statistic. According to models of tournament entry (Niederle and Vesterlund, 2007; Cotton et al., 2014; Lazear et al., 2018) and directed search (Wright et al., 2019; Belot et al., 2018), we should expect low ability people not to apply when receiving information on an outstanding past performance vis-á-vis moderate past performance. This is because low-skilled people would naturally shun away from increased competition or a more selective employer. This would imply, consequently, an increase in average quality in the treatment featuring the 89% statistic, which is in contrast with the evidence shown in the paper.[69]

9.2.2 On-the-job dating market

Candidates may interpret employer's recruitment messages in terms of dating opportunities. Under this hypothesis, we expect the reaction to photographs to differ by sexuality and marital status. First, we expect heterosexual (non-heterosexual) men to apply more when seen a female (male) photograph. Secondly, the positive effects of seeing a person of the opposite gender should be weaker for married people. Table A.8 tests for differential treatment effects of the male photograph on applications by sexual orientation (in odd Columns) and by marital status (in even Columns). Overall, I do not find support for the on-the-job dating channel. First, there are no significant differences in the effect of the treatment based on marital status. Secondly, non-heterosexual men and women both react positively to the male photograph, suggesting that their reaction cannot be about dating.

9.2.3 Gender differences in preferences

Different variance in past success among workers could affect the perceived riskiness of performance in the job, leading to different reactions by gender as men tend to be less risk averse than women (Holt and Laury, 2002; Dohmen et al., 2005; Eckel and Grossman, 2008). There are two main comments against this interpretation. First, more variance in past success does not necessarily imply higher uncertainty

[68] Team assignment is orthogonal to expected performance and based on candidate's regional preferences and diversity considerations of the partner organization.

[69] Moreover, the organization has a strong reputation for being a selective employer, which anecdotally was not challenged by the information given in two treatment groups.

if people know their own ability. A high aggregate past performance could even be associated with higher uncertainty if people think that success is determined by other (unclear) factors rather than ability. Secondly, even allowing higher expected returns to ability to be associated with higher risk, we should expect women to apply less in this condition on average.

Another stream of work shows that men tend to be more overconfident than women (Lundeberg et al., 1994; Beyer and Bowden, 1997; Niederle and Vesterlund, 2007; Grosse et al., 2014; Dreber et al., 2014; Coffman, 2014), but the gender gap shrinks or even reverses in typical female domains (Coffman et al., 2019). In line with these results, Table A.9 shows that men in my sample tend to be less overconfident than women, especially in job-specific skills. Appendix F further shows that the increase in men's applications is driven by men with low confidence in their estimates of people's performance in female-dominated jobs. As long as confidence about others' performance is correlated with confidence in one's own ability, it suggests that the effects are actually driven by the least confident men (Moore and Healy, 2008).

Finally, high returns to ability might signal that the job is competitive. Well-known results are that men are more likely to select into competitive environments than women (Niederle and Vesterlund, 2007) and that this gap is larger for tasks which are perceived as more "masculine" (Dreber et al., 2014; Grosse et al., 2014; Flory et al., 2015).[70] This interpretation presupposes that beliefs about the returns to ability must have changed, otherwise people would have no reason to become competitive. Thus my main interpretation of the information treatment is still needed for preferences for competition to explain the results. I also check whether reaction to the treatment differ by participants' competitive background. I identify two types of candidates: those used to competition, who studied a male-dominated subject in a top-tier university, and those not used to competition, who studied a female-dominated subject in lower-tier universities.[71] Figure A.7 shows that both men and women react similarly to the expectations treatment independently of this proxy of competitiveness, suggesting that competitive attitudes are not the driving force of the results.

9.2.4 Attention

Photographs may differ in the extent to which they capture the agent's attention (Gabaix, 2019; Mas and Pallais, 2017), which in turn can affect their decision to apply. I use requests for reminders of the "unique candidate number" as a proxy for endogenous (in)attention to the intervention. The idea is that candidates who asked for fewer reminders have either paid more attention to the original invitation-to-apply email, where the number was reported, or have looked back at it several times.

Table A.7 shows that men pay relatively more attention to the female than the male photograph, which contributes to explaining the slightly higher application rate by men in the female-photograph treatment. This is in contrast with models of salience (Bordalo et al., 2013) and attention triggered by perceived similarity (Forehand and Deshpandé, 2001), but suggests that visceral influences might play a role in explaining job applications Loewenstein (1996); Bertrand et al. (2010).

[70] The intervention did not change the structure of incentives on the job, as in Flory et al. (2015) and people know their earnings do not depend on rankings. Moreover, sorting in female-dominated jobs is usually negatively correlated with competitiveness Buser et al. (2014).

[71] The performance of these two groups once on the job is the same on average.

10 Concluding remarks

Blue-collar employment is shrinking across the developed world (David et al., 2013). These trends challenge the traditional role of men both in society and within households by creating male idleness and financial insecurity, especially on the left tail of the ability distribution (Dorn et al., 2018; Coile and Duggan, 2019). At the same time, female-dominated sectors such as healthcare and education are growing and face relatively little risk of automation in the future (Nedelkoska and Quintini, 2018). And yet, male labour supply is still relatively untapped as a resource for addressing the shortage of teachers and nurses in many industrialized economies. Understanding the interaction between traditional gender norms and gender-specific information in rapidly changing labour markets is crucial to allowing men in declining industries achieve new opportunities (Binder and Bound, 2019).

In this paper, I provided evidence that the limited entry of men into female-dominated jobs can be explained by limited information on returns to ability rather than job-gender composition. I show that providing information on the chances of standing out increases men's applications, especially when their experience in the sector is limited. Men with expectations of higher returns to ability are more likely to be hired, perform better and are happier on the job. At the same time, a higher male share discourages the entry and retention of less talented women.

Historically, job advertisement has been a common strategy to change the demographic composition of male-dominated occupations. Rosie the Riveter is a long-lived testimonial of the crucial role of advertising in recruiting women to supply-short male jobs during WWII (Honey, 1985; Milkman, 1987). This legacy inspired recent attempts to attract men to female-dominated sectors portraying masculine men working as nurses or teachers. My results are a cautionary tale against strategies designed to promote a new male identity in these roles without addressing informational constraints.

Informational asymmetries between men and women are the central force in this paper, which assumes that men and women only differ in terms of their informational endowments. This assumption contrasts with many studies on gender differences in preferences (for a review see Bertrand, 2011) and moves the focus of research from natural to nurturing differences, which emerge as a result of being the minority in a certain occupation. However, more research is needed to understand whether informational constraints interact with other types of expectations that might differ between the minority and the majority.

Many other questions are left for future research. How do informational asymmetries between men and women form? How do supply-side and demand-side factors interact in determining whether men apply and whether they get hired in female-dominated jobs? Hopefully answering these questions may prevent communities such as the ones in the Rust Belt or the North of England from being left behind by a rapidly evolving economy.

References

Abebe, G., Caria, S., and Ortiz-Ospina, E. (2017). The selection of talent experimental and structural evidence from ethiopia.

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Akerlof, G. A. and Kranton, R. E. (2005). Identity and the economics of organizations. *Journal of Economic Perspectives*, 19(1):9–32.

Al-Ubaydli, O., List, J. A., and Suskind, D. L. (2017). What can we learn from experiments? understanding the threats to the scalability of experimental results. *American Economic Review*, 107(5):282–86.

Alan, S., Ertac, S., and Mumcu, I. (2018). Gender stereotypes in the classroom and effects on achievement. *Review of Economics and Statistics*, 100(5):876–890.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence. *Journal of Political Economy*, 97(6):1447–1458.

Arndt, M. and Bigelow, B. (2005). Professionalizing and masculinizing a female occupation: The reconceptualization of hospital administration in the early 1900s. *Administrative Science Quarterly*, 50(2):233–261.

Arrow, K. (1973). The theory of discrimination. *Discrimination in Labor Markets*, 3(10):3–33.

Arrow, K. J. (1998). What has economics to say about racial discrimination? *Journal of Economic Perspectives*, 12(2):91–100.

Ashraf, N., Bandiera, O., Davenport, E., and Lee, S. S. (2020). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. *American Economic Review*, 110(5):1355–94.

Baranov, V., De Haas, R., and Grosjean, P. A. (2020). Men. roots and consequences of masculinity norms. *Roots and Consequences of Masculinity Norms (March 11, 2020). UNSW Business School Research Paper.*

Becker, G. (1957). The economics of discrimination. *Chicago: University of Chicago Press.*

Belot, M., Kircher, P., and Muller, P. (2018). How wage announcements affect job search: A field experiment. *IZA Working Papers*, (11814).

Benjamin, D. J., Choi, J. J., and Strickland, A. J. (2010). Social identity and preferences. *American Economic Review*, 100(4):1913–28.

Benyishay, A., Jones, M., Kondylis, F., and Mushfiq, A. (2016). *Are Gender Differences in Performance Innate or Socially Mediated?* The World Bank.

Bertrand, M. (2011). New perspectives on gender. In *Handbook of Labor Economics*, volume 4, pages 1543–1590. Elsevier.

Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, 95(2):94–98.

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. In *Handbook of Economic Field Experiments*, volume 1, pages 309–393. Elsevier.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., and Zinman, J. (2010). What's advertising content worth? evidence from a consumer credit marketing field experiment. *The Quarterly Journal of Economics*, 125(1):263–306.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013.

Besley, T. and Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review*, 95(3):616–636.

Beyer, S. and Bowden, E. M. (1997). Gender differences in self-perception: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, 23(2):157–172.

Binder, A. J. and Bound, J. (2019). The declining labor market prospects of less-educated men. *Journal of Economic Perspectives*, 33(2):163–190.

Blau, F. D., Ferber, M. A., and Winkler, A. E. (2013). *The Economics of Women, Men and Work.* Pearson Higher Ed.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

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

Boneva, T., Rauh, C., et al. (2017). Socio-economic gaps in university enrollment: The role of perceived pecuniary and non-pecuniary returns. Technical report, CESifo Group Munich.

Booth, A. and Leigh, A. (2010). Do employers discriminate by gender? a field experiment in female-dominated occupations. *Economics Letters*, 107(2):236–238.

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

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–73.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Breda, T., Grenet, J., Monnet, M., and Van Effenterre, C. (2020). Do female role models reduce the gender gap in science? evidence from french high schools. Technical report, Institute of Labor Economics (IZA).

Brewer, M. B. and Hewstone, M. E. (2004). *Self and Social Identity.* Blackwell publishing.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, 110(10):2997–3029.

Bursztyn, L. and Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9:131–153.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447.

Cameron, C. (2001). Promise or problem? a review of the literature on men working in early childhood services. *Gender, Work & Organization*, 8(4):430–453.

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

Charles, K. K., Guryan, J., and Pan, J. (2018). The effects of sexism on american women: The role of norms vs. discrimination. *NBER Working Paper*, (w24904).

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660.

Coffman, K. B., Collis, M., and Kulkarni, L. (2019). *Stereotypes and Belief Updating.* Number 19-068.

Coffman, L. C., Featherstone, C. R., and Kessler, J. B. (2017). Can social information affect what job you choose and keep? *American Economic Journal: Applied Economics*, 9(1):96–117.

Coile, C. C. and Duggan, M. G. (2019). When labor's lost: Health, family life, incarceration, and education in a time of declining economic opportunity for low-skilled men. *Journal of Economic Perspectives*, 33(2):191–210.

Cortes, P. and Pan, J. (2018). Occupation and gender. *The Oxford Handbook of Women and the Economy*, pages 425–452.

Cotton, C., Hickman, B. R., and Price, J. (2014). Affirmative action and human capital investment: Evidence from a randomized field experiment. *NBER Working Paper*, (w20397).

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.

Dal Bó, E., Dal Bó, P., and Eyster, E. (2017). The demand for bad policy when voters underappreciate equilibrium effects. *The Review of Economic Studies*, 85(2):964–998.

Dal Bó, E., Finan, F., and Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, 128(3):1169–1218.

David, H., Dorn, D., and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–68.

Del Carpio, L. and Guadalupe, M. (2018). More women in tech? evidence from a field experiment addressing social identity. *Working Paper*, (13234).

Deserranno, E. (2019). Financial incentives as signals: Experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics*, 11(1):277–317.

Dohmen, T. J., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey. *Working Paper*, (1730).

Dorn, D., Hanson, G., et al. (2018). When work disappears: Manufacturing decline and the falling marriage market value of young men. *American Economic Review: Insights*.

Dreber, A., von Essen, E., and Ranehill, E. (2014). Gender and competition in adolescence: Task matters. *Experimental Economics*, 17(1):154–172.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American sociological review*, 20(2):210–217.

Dweck, C. S. and Elliot, A. J. (2005). *Handbook of Competence and Motivation*. Guilford Press New York.

Eckel, C. C. and Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, 1:1061–1073.

Eden, D. (1992). Leadership and expectations: Pygmalion effects and other self-fulfilling prophecies in organizations. *The Leadership Quarterly*, 3(4):271–305.

Eisenberger, N. I., Lieberman, M. D., and Williams, K. D. (2003). Does rejection hurt? an fmri study of social exclusion. *Science*, 302(5643):290–292.

England, P. (2017). *Households, employment, and gender: A social, economic, and demographic view*. Routledge.

Ensmenger, N. (2012). *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*. History of Computing. MIT Press.

Fisman, R., Iyengar, S. S., Kamenica, E., and Simonson, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, 121(2):673–697.

Flory, J. A., Leibbrandt, A., and List, J. A. (2015). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies*, 82(1):122–155.

Flory, J. A., Leibbrandt, A., Rott, C., and Stoddard, O. (2019). Increasing workplace diversity: Evidence from a recruiting experiment at a fortune 500 company. *Journal of Human Resources*, pages 0518–9489R1.

Flynn, S. (2006). Campaign to attract more men into teaching. *https://www.irishtimes.com/news/campaign-to-attract-more-men-into-teaching-1.1005448*.

Folke, O. and Rickne, J. (2020). Sexual harassment and gender inequality in the labor market. Technical report, CEPR Discussion Papers.

Forehand, M. R. and Deshpandé, R. (2001). What we see makes us who we are: Priming ethnic self-awareness and advertising response. *Journal of Marketing Research*, 38(3):336–348.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 261–343. Elsevier.

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from french grocery stores. *The Quarterly Journal of Economics*, 132(3):1219–1260.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American Economic Review*, 96(2):1–21.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4):1091–1119.

Gonzalez, L. and Loureiro, Y. K. (2014). When can a photo increase credit? the impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance*, 2:44–58.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6):1464.

Grosse, N., Riener, G., and Dertwinkel-Kalt, M. (2014). Explaining gender differences in competitiveness: Testing a theory on gender-task stereotypes. *Working Paper*, (017).

Haile, G. A. (2012). Unhappy working with men? workplace gender diversity and job-related well-being in britain. *Labour Economics*, 19(3):329–350.

Handel, B. and Schwartzstein, J. (2018). Frictions or mental gaps: What's behind the information we (don't) use and when do we care? *Journal of Economic Perspectives*, 32(1):155–78.

Hoff, K. and Pandey, P. (2006). Discrimination, social identity, and durable inequalities. *American Economic Review*, 96(2):206–211.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655.

Honey, M. (1985). Creating rosie the riveter: Class, gender, and propaganda during world war ii. *Minerva*, 3(2):65.

Hoogendoorn, S. and Van Praag, C. (2012). Ethnic diversity and team performance: a randomized field experiment. In *Academy of Management Proceedings*, volume 1, page 13736. Citeseer.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and u.s. economic growth.

Econometrica, 87(5):1439–1474.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548.

Katz, L. F. (2014). America's jobs challenges and the continuing role of the us department of labor. *ILR Review*, 67(3_suppl):578–583.

Lazear, E. P. (1998). Diversity and immigration. *NBER Working Paper*, (w6535).

Lazear, E. P., Shaw, K. L., and Stanton, C. T. (2018). Who gets hired? the importance of competition among applicants. *Journal of Labor Economics*, 36(S1):S133–S181.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174.

Lin, V. W., Lin, J., and Zhang, X. (2015). Us social worker workforce report card: Forecasting nationwide shortages. *Social Work*, 61(1):7–15.

List, J. A. and Metcalfe, R. (2014). Field experiments in the developed world: An introduction. *Oxford Review of Economic Policy*, 30(4):585–596.

List, J. A. and Rasul, I. (2011). Field experiments in labor economics. In *Handbook of Labor Economics*, volume 4, pages 103–228. Elsevier.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65(3):272–292.

Lordan, G. and Pischke, J.-S. (2016). Does rosie like riveting? male and female occupational choices. Technical report, London School of Economics and Political Science, LSE Library.

Lundeberg, M. A., Fox, P. W., and Punćcohaŕ, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, 86(1):114–21.

Marinescu, I. and Wolthoff, R. (2020). Opening the black box of the matching function: The power of words. *Journal of Labor Economics*, 38(2):535–568.

Marx, D. M. and Ko, S. J. (2012). Superstars "like" me: The effect of role model similarity on performance under threat. *European Journal of Social Psychology*, 42(7):807–812.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–3759.

McGonagle, E. (2019). Nhs celebrates male nurses in mission to overcome gender divide. *https://www.campaignlive.co.uk/article/nhs-celebrates-male-nurses-mission-overcome-gender-divide/1661044*.

Milkman, R. (1987). *Gender at Work: The Dynamics of Job Segregation by Sex During World War II*, volume 308. University of Illinois Press.

Miller, C. C. (2017). Why men don't want the jobs done mostly by women. *The New York Times*.

Miller, T. and del Carmen Triana, M. (2009). Demographic diversity in the boardroom: Mediators of the board diversity–firm performance relationship. *Journal of Management studies*, 46(5):755–786.

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2):502.

Nedelkoska, L. and Quintini, G. (2018). Automation, skills use and training. *OECD Social, Employment, and Migration Working Papers*, (202):0_1–119.

News, N. (2019). Young male nursing applicants surge after 'we are the nhs' recruitment campaign.

Ngai, L. R. and Petrongolo, B. (2017). Gender gaps and the rise of the service economy. *American Economic Journal: Macroeconomics*, 9(4):1–44.

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from madagascar. *Unpublished manuscript*.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Niederle, M. and Yestrumskas, A. H. (2008). Gender differences in seeking challenges: The role of institutions. Technical report, National Bureau of Economic Research.

OECD (2019). Oecd employment outlook 2019 the future of work. *OECD Publishing, Paris.*

Oh, S. (2019). Does identity affect labor supply? *Job Marker Paper, Columbia University.*

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, 26(3):291–310.

Pan, J. (2015). Gender segregation in occupations: The role of tipping and social interactions. *Journal of Labor Economics*, 33(2):365–408.

Payne, B. D. (1984). The relationship of test anxiety and answer-changing behavior: An analysis by race and sex. *Measurement and Evaluation in Guidance*, 16(4):205–210.

Perkins, L. A., Thomas, K. M., and Taylor, G. A. (2000). Advertising and recruitment: Marketing to minorities. *Psychology & Marketing*, 17(3):235–255.

Porter, C. and Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 12(3):226–54.

Ravalier, J. M. (2019). Psycho-social working conditions and stress in uk social workers. *The British Journal of Social Work*, 49(2):371–390.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186.

Rich, J. (2014). What do field experiments of discrimination in markets tell us? a meta analysis of studies conducted since 2000. *Working Paper*, (8584).

Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science*, 3(6):176–179.

Rudman, L. A. and Fairchild, K. (2004). Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, 87(2):157.

Samek, A. (2019). Gender differences in job entry decisions: A university-wide field experiment. *Management Science*, 65(7):3272–3281.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., and Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social psychology*, 61(2):195.

Steele, C. M. and Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, 69(5):797.

Stinebrickner, R. and Stinebrickner, T. R. (2014). A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81(1):426–472.

Stone, J., Perry, W., and Darley, J. M. (1997). "white men can't jump": Evidence for the perceptual confirmation of racial stereotypes following a basketball game. *Basic and Applied Social Psychology*, 19(3):291–306.

Tajfel, H. and Turner, J. (1986). The social identity theory of intergroup behavior. *Psychology of Intergroup Relations*, 5:7–24.

The Economist (2018). A shortage of nurses calls for the recruiting of more men. but that is a tricky task. *The Economist*.

Tipper, J. (2004). How to increase diversity through your recruitment practices. *Industrial and Commercial Training*, 36(4):158–161.

Whittingham, A. (2018). Government must act to halt crisis in social care recruitment. *The Guardian*.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *Review of Economic Studies*, 82(2):791–824.

Wiswall, M. and Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507.

Wozniak, D. et al. (2009). Choices about competition: Differences by gender and hormonal fluctuations, and the role of relative performance feedback. Technical report, University Library of Munich, Germany.

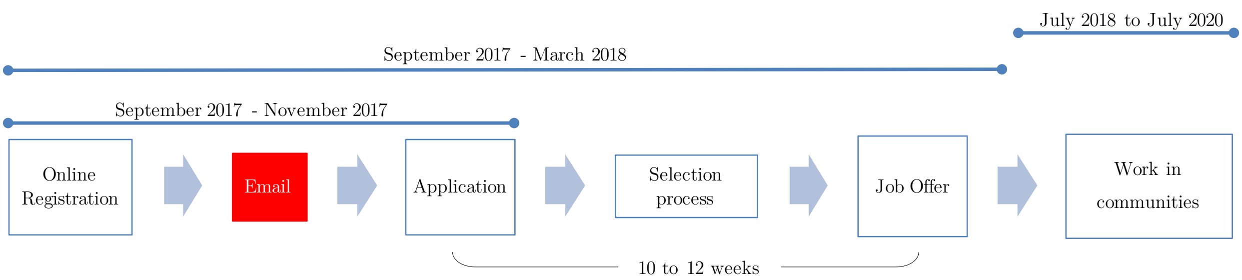
Wright, R., Kircher, P., Julien, B., and Guerrieri, V. (2019). Directed search and competitive search: A guided tour. *Journal of Economic Literature*.

Young, A. (2018). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2):557–598.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595.

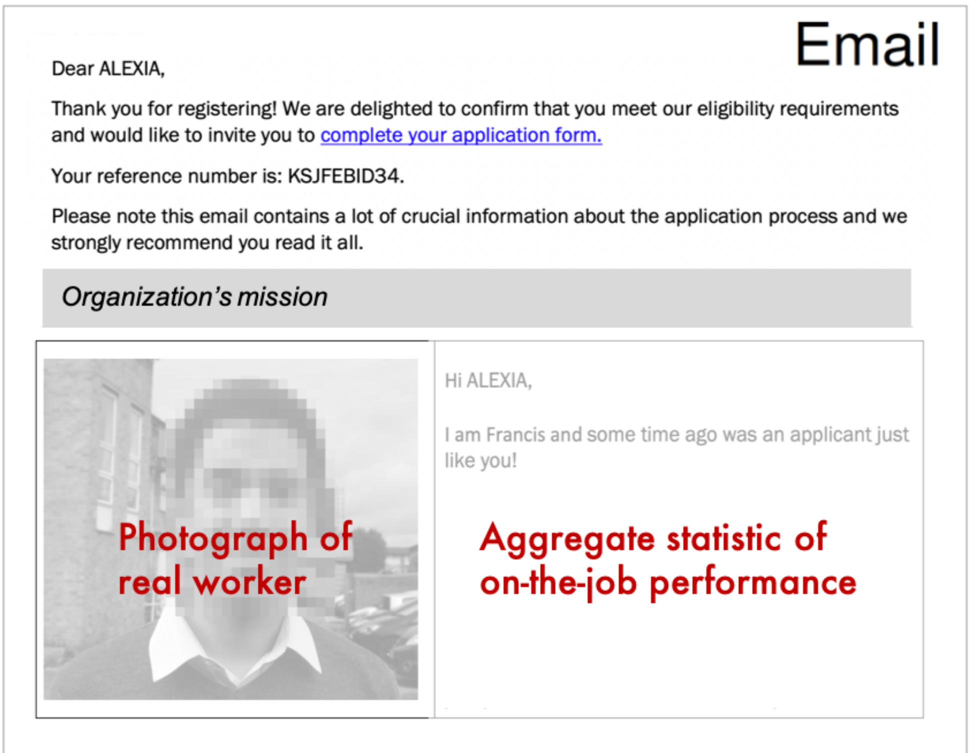
11 Figures

Figure 1. Recruitment timeline



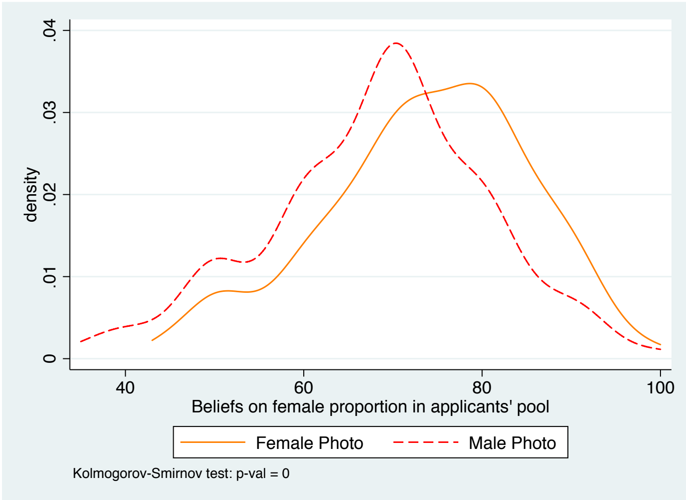
Note. The Figure shows the recruitment timeline of the partner organization. Applications were open from September until November 2017. A given candidate was randomized in one of the different invitation-to-apply emails between his/her online registration and application. The organization completed the recruitment process in March 2018. For a successful applicant, it usually took between ten and twelve weeks between application and job offer. If a person was hired and accepted the job, actual work in local communities started in July 2018. Performance data are collected for the whole duration of the two-year programme between July 2018 and July 2020.

Figure 2. Intervention email template



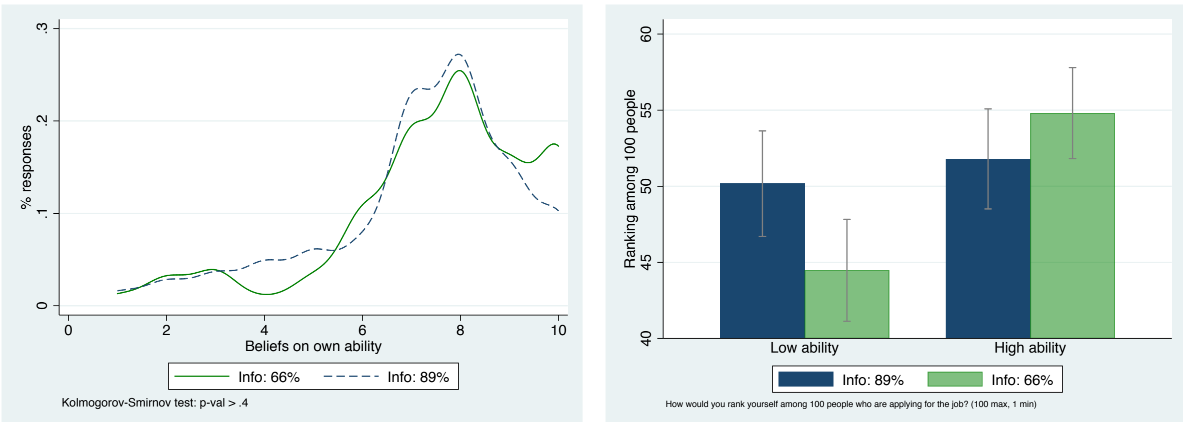
Note. The Figure shows a stylized example of one of the email templates used in the intervention.

Figure 3. Gender shares shock: manipulation checks



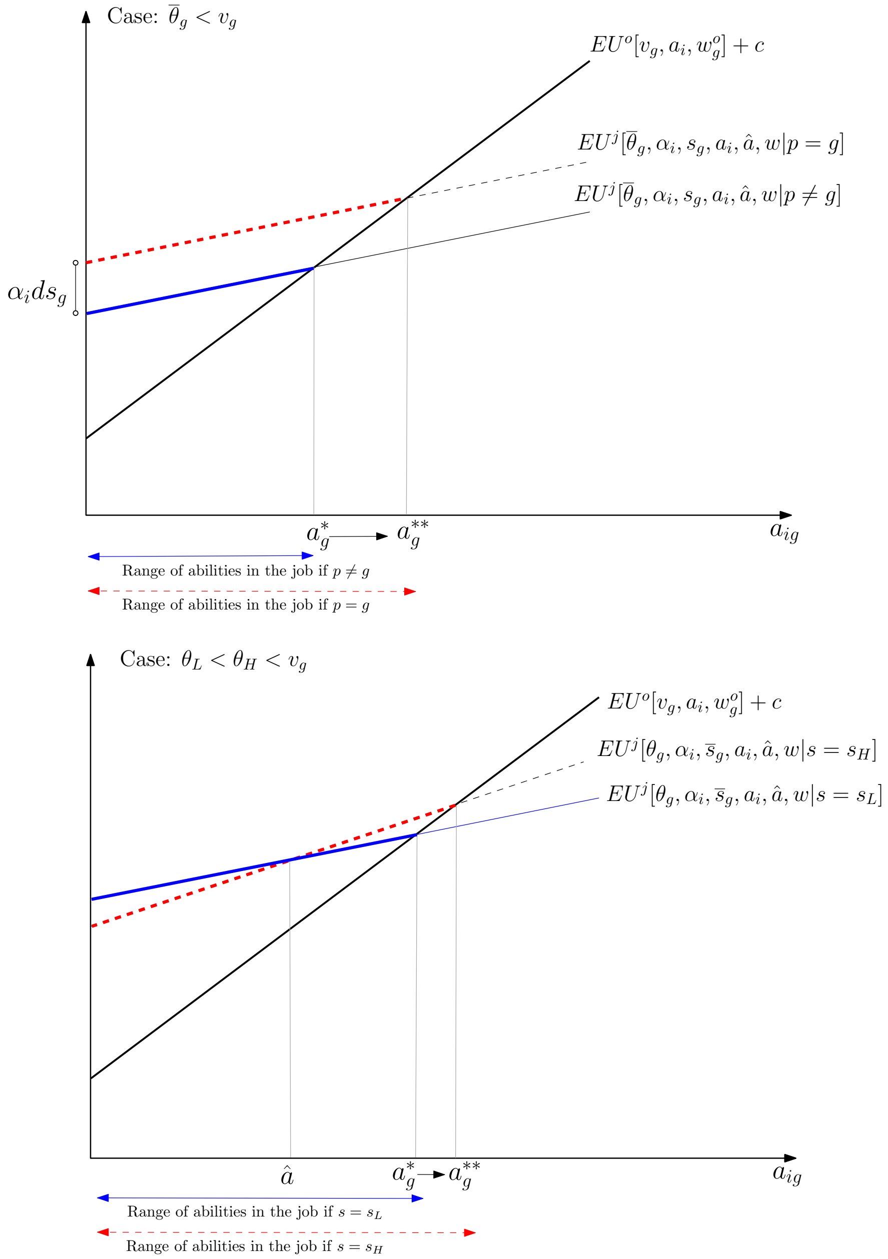
Note. The figure shows the distribution of answers to the question "Consider 100 people who apply for this job. How many do you think are women?", separately for respondents assigned to the email with a female or male photograph. Data are from the auxiliary online surveys. The dashed (solid) line is for the male (female) photograph treatment. The number of respondents is 504, of whom 262 are from the Prolific Academic sample and 242 from the organization's sample.

Figure 4. Expectations shock: manipulation checks



Note. The left panel shows the distribution of answers to the question "How do you expect a person with your skills and experience to perform in interacting with families in need?" on a scale from 1 (min) to 10 (max), separately for respondents assigned to the email with a statistic of 66% (solid line) or 89%(dashed line) of past high achievers. The right panel shows mean answers to the question "Consider 100 people who are applying for this job. Based on the ad you just viewed, on a scale from 100 (best) to 1 (worst), how would you rank yourself for the job among them?", by information treatment and ability level. The ability level is defined above or below the median of the answers reported in the left-hand side graph. Green bars are for the 66% statistic and blue bars for the 89% statistic. Data are from the auxiliary online surveys. The number of respondents is 504, of whom 262 are from the Prolific Academic sample and 242 from the organization's sample.

Figure 5. Theory predictions: shock to perceived gender shares and expectations of returns to ability in the case of negative sorting in social work



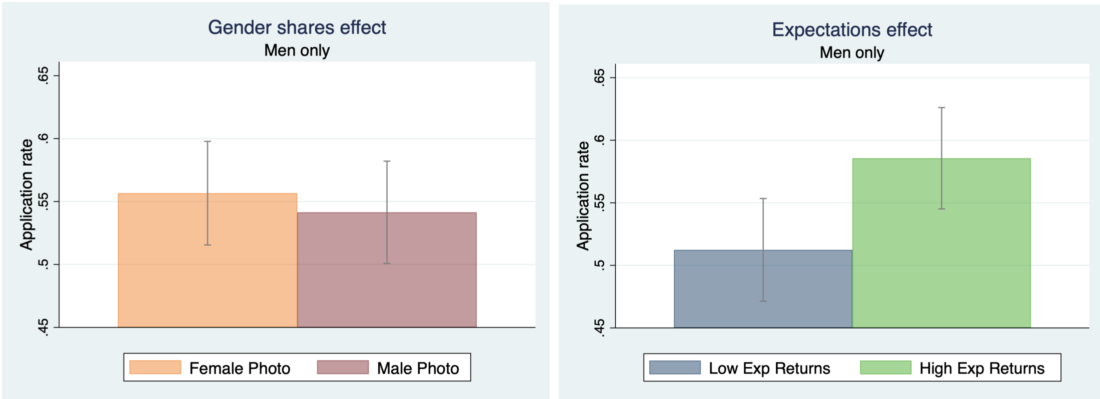
Note. The figure plots the application decision for potential applicants of gender g. Both panels consider the case $U^{j'}(a_i) < U^{o'}(a_i)$. The top panel shows the effect of a shock to perceived gender shares and the bottom panel to expectations of returns to ability. The solid thick line shows expected utility in the outside option. The case $U^{j'}(a_i) > U^{o'}(a_i)$ is described in the Appendix.

Top panel: the dashed and thin solid lines show the expected job utility when receiving a photo of the same ($p = g$) or different gender ($p \neq g$), respectively. The vertical distance between these two lines comes from the assumption of the model $E[s_g|p = g] > E[s_g|p \neq g]$. The two thresholds of ability for the marginal applicants a_g^* and a_g^{**} are determined from the intersection of the expected job utility and expected outside option. From Result 1, the size of the applicants' pool is greater when $p = g$ than $p \neq g$ and the marginal applicant a_g^{**} is more skilled than a_g^*.

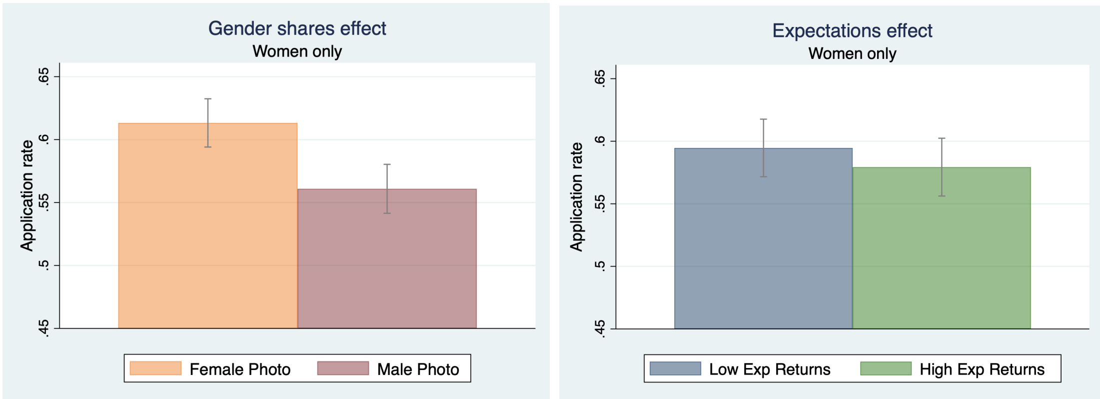
Bottom panel: the two thin dashed and solid lines show the expected job utility when receiving information on high ($s = s_H$) or low ($s = s_L$) returns to ability. The different slope of these two lines is explained by $E[\theta|s = s_H] > E[\theta|s = s_L]$, as higher returns to ability correspond to a higher slope. The two thresholds of ability for the marginal applicants a_g^* and a_g^{**} are determined from the intersection of the expected job utility and expected outside option. From Result 2, the applicants' pool is larger when $s = s_H$ than $s = s_L$ as long as $B > 0$ and $U^{j'}(a_i) < U^{o'}(a_i)$ and the marginal applicant a_g^{**} is more skilled than a_g^* if $B > 0$.

Figure 6. Application rates by treatment and gender

(a) Men



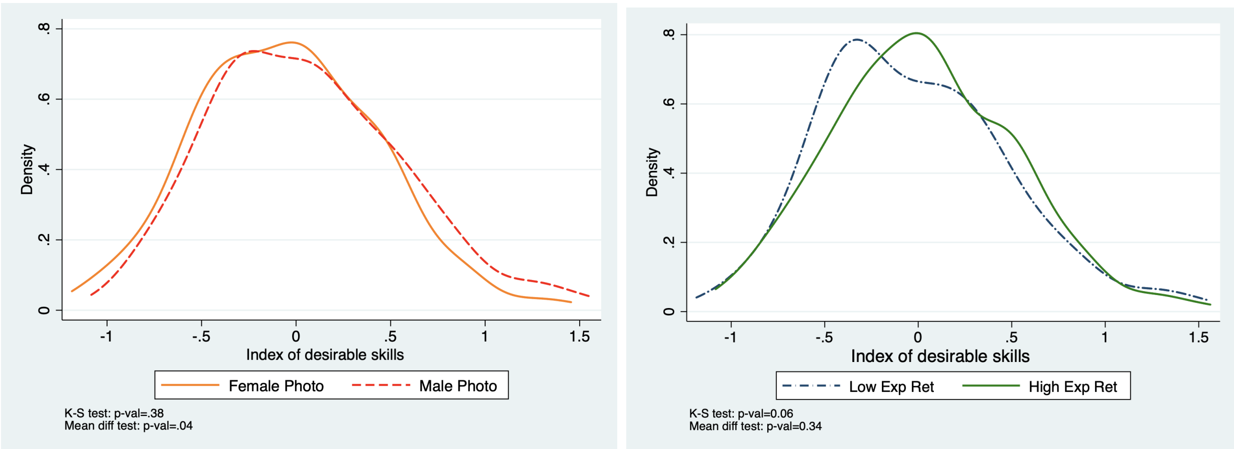
(b) Women



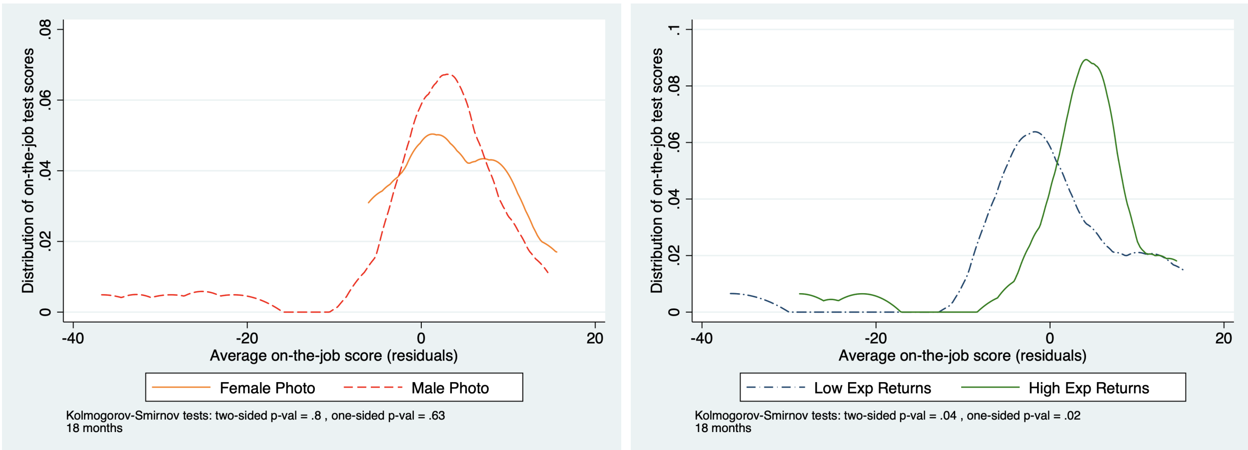
Note. Panel A shows application rates for men by photograph treatment (left-hand side) and information treatment (right-hand side). Panel B shows application rates for women by photograph treatment (left-hand side) and information treatment (right-hand side). Error bars show 95% confidence intervals.

Figure 7. Male applicants' qualifications and average on-the-job test scores by treatment

(a) Qualifications

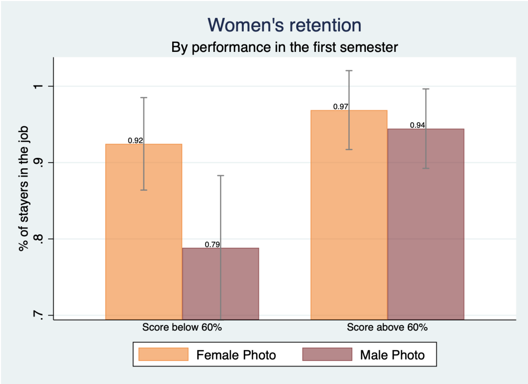


(b) Average on-the-job test scores



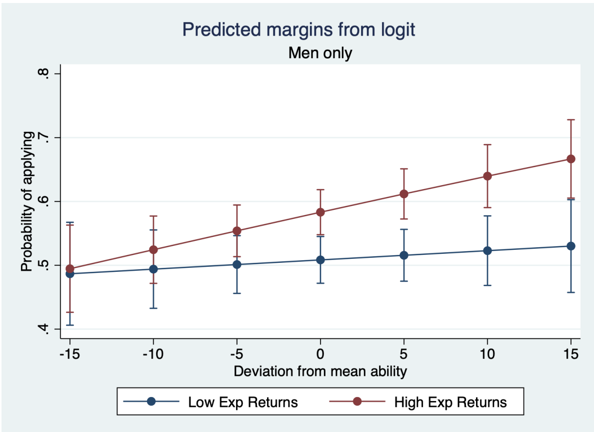
Note. The figure shows the distribution of male applicants' observable skills (a) and the distribution of men's average on-the-job test scores (b). The index in Figure (a) is the "desirable skills index" computed as the mean of the following standardized variables: receiving a first grade, being from a top tier university, frequent past volunteering, high cognitive skills and score in English pre-university tests. The average on-the-job test score in Figure (b) is the weighted average of the scores in the nine performance assessments required in the job, where weights are given by the credits assigned to each test by the organization and residualized after controlling for ethnicity, past application and early registration. Figures on the left-hand side show the distributions by photograph treatment and the dashed lines are for the male photograph. Figures on the right-hand side show the distributions by information treatment and the dashed lines are for high expected returns to ability.

Figure 8. Women's retention by performance



Note. The figure shows female workers' retention by their performance on the job. The figure shows the retention rate for women by photograph treatment and by the level of performance in the first semester on the job. I categorize on-the-job performance into two levels: above or below 60% average score in the first semester. Retention is defined as being still in the programme by the end of the second year.

Figure 9. Predicted application probability: logit estimation



Note. The figure shows predictive margins from the logit discrete choice model for men only. The variable on the x-axis is the de-meaned predicted on-the-job performance. Predicted on-the-job performance is calculated using a truncated linear regression on the pure control group only with the following independent variables: ranking and average completion rate of the university attended by the candidate, subject studied, obtaining a first grade, whether the grade is expected or obtained, age, age squared and whether the person is in FTE.

12 Tables

Table 1. Balance checks and summary statistics

	Men			Women			Joint		Pairwise
VARIABLES	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*F-stat*	*p-val*	*min p-val*
Demographics									
Male	1013	1.00	0.00	4404	0.00	0.00	0.04	1.0	0.72
Non-white	1013	0.28	0.45	4404	0.27	0.45	0.08	1.0	0.60
Age	1013	28.7	9.2	4404	26.4	7.9	0.29	0.88	0.42
Married	995	0.19	0.4	4331	0.12	0.33	0.19	0.95	0.47
Caring duties	1013	0.16	0.36	4404	0.16	0.37	0.96	0.43	0.11
Non heterosexual	959	0.13	0.34	4131	0.07	0.26	0.36	0.84	0.33
Education and employment									
Top UK university	1013	0.33	0.47	4404	0.32	0.47	0.205	0.936	0.38
First grade	1013	0.2	0.4	4404	0.18	0.39	0.697	0.594	0.13
Graduate	1013	0.46	0.5	4404	0.35	0.48	0.473	0.756	0.19
Scientific subject	1013	0.09	0.28	4404	0.05	0.21	0.496	0.738	0.18
FTE	1013	0.49	0.5	4404	0.42	0.49	0.25	0.911	0.41
in: public sector	500	0.46	0.5	1840	0.56	0.5	1.06	0.373	**0.05**
in: healthcare	500	0.16	0.36	1840	0.17	0.37	0.87	0.483	0.11
in: corporate/business	500	0.32	0.47	1840	.22	.41	1.17	0.324	**0.05**
Registration									
Past application	1013	0.07	0.26	4404	0.06	0.24	0.08	0.99	0.61
Pre-submission call	1013	0.11	0.32	4404	0.08	0.28	0.48	0.75	0.27
Early registration	1013	0.04	0.2	4404	0.05	0.21	0.31	0.87	0.40
Registration before November	1013	0.53	0.5	4404	0.57	0.5	0.02	1.00	0.83
Any event	1013	0.00	0.05	4404	0.01	0.11	0.13	0.97	0.52
Socio-economic background									
Economic school support	1013	0.27	0.44	4404	0.27	0.45	0.62	0.65	0.15
Low socio-econ status	1013	0.60	0.49	4404	0.62	0.49	1.23	0.30	**0.08**
Young carer	999	.04	.2	4339	.04	.2	0.62	0.15	**0.02**
Care leaver	1006	.03	.17	4369	.02	.15	0.46	0.76	0.26

Note. The Table shows summary statistics for the overall experimental sample. "Caring Duties" is a dummy equal to one if the respondent is a primary or secondary carer of children. I define top UK universities those belonging to the Russell Group (http://russellgroup.ac.uk). "First Grade" is a dummy for achieving an average of A across classes. "Graduate" indicates whether the candidate graduated in 2016 or before. "Scientific Subject" is equal to one if the person studied Engineering, IT/Computer Science, Maths or Natural Sciences. "Past application" is equal to one if the candidate applied already in the past for the same job. "Pre-submission call" indicates whether the candidate received a call from a recruitment officer to encourage application. "Early registration" is a dummy equal to one if the person had access to an early opening of the application. "Registration before November" is a dummy for whether the person registered online before the 1^{st} of November. "Any event" is a dummy equal to one if the candidate participated in any of the organization's career events. "Economic school support" equals one if the candidate received free school meals or any other type of economic support (e.g., scholarship) during school. "Low socio-econ status" equals one if the occupation of the household's highest earner in candidate's family was unemployment, routine manual or routine semi-manual or for parents with no degree. "Care leaver" is a dummy equal to one if the person spent some time looked after by a social worker before the age of 14. Columns 4 and 5 (under "Joint") report the F-statistic and p-value from a joint test of the significance of the set of treatment dummies in explaining each variable in a regression with pooled genders and with robust standard errors. The last Column reports the minimum p-value from the associated t-test between pairs of treatment groups with robust standard errors and with pooled genders. I also fail to reject the null hypothesis of zero effect of all the variables reported in the Table in a joint test of orthogonality on assignment to any treatment group (F(23, 4865)=0.67).

Table 2. Men's results

	(1)	(2)	(3)	(4)	(5)	(6)
	Application and selection process			On-the-job outcomes		
VARIABLES	Applied and never DO	Received Offer	Accepted Offer	Mean on-the-job score Semester 1	Semesters 2/3	Early Exit
Male Photo	-0.017	0.055	0.090	-3.699	-6.122	0.059
	(0.035)	(0.034)	(0.124)	(2.669)	(4.962)	(0.092)
High Exp Returns	0.071**	0.061*	-0.023	5.479*	4.183	-0.072
	(0.035)	(0.033)	(0.128)	(2.917)	(5.684)	(0.106)
Observations	807	440	67	43	43	43
R-squared	0.018	0.062	0.035	0.210	0.104	0.076
Basic Controls	✓	✓	✓	✓	✓	✓
Mean Dep Var	0.52	0.10	0.70	58.07	53.27	0.14
Mean Dep Var in Pure C	0.53	0.21	0.83	56.16	45.42	0.32
Photo = Exp Ret p-val	0.08	0.89	0.53	0.02	0.15	0.27
Rand Inf p-val						
Photo	0.63	0.11	0.47	0.16	0.25	0.51
Exp Returns	0.04	0.08	0.83	0.06	0.47	0.47

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates for men only. The table reports results of six different regressions. The omitted category is the treatment group which received the female photograph and information on low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high returns to ability (specification (2) of Section 5.1). The dependent variables are indicator dummies for application, receiving a job offer (conditional on applying) and accepting the job offer (conditional on receiving the offer) in Columns (1), (2) and (3). The dependent variable in column (4) is the average on-the-job test score achieved in the first five assessments during the first semester on the job and in column (5) is the average on-the-job test score achieved in the four additional assessments during the second and third semester on the job. The score goes from 0 to 100 in each test, the average is weighted by the credits assigned to each exam by the organization. The dependent variable in column (6) is a dummy equal to one if the person left the programme before completing it. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration, non-white ethnicity. The rows "Rand Inf p-val" contain the p-values of the coefficients on the indicated treatment dummies from randomization inference (randomization-t) with 1000 repetitions.

Table 3. New male hires: on-the-job performance

	DV: On-the-Job Std. Scores			
	(1)	(2)	(3)	(4)
Male Photo	-0.284	-0.300*	-0.293	-0.305*
	(0.194)	(0.166)	(0.194)	(0.168)
High Exp Returns	0.248*	0.225	0.258*	0.234
	(0.141)	(0.144)	(0.143)	(0.147)
Observations	387	387	387	387
R-squared	0.193	0.218	0.202	0.223
Basic Controls	✓	✓	✓	✓
Exam FE	✓	✓	✓	✓
Quality Controls	×	✓	×	✓
Location Difficulty Controls	×	×	✓	✓
Mean Dep Var	0.19	0.19	0.19	0.19
Mean Dep Var in Pure C	-0.12	-0.12	-0.12	-0.12
Photo = Exp Ret p-val	0.03	0.03	0.03	0.03

Clustered s.e. in parentheses (worker level)

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS panel estimates for men only. The table reports results of four different regressions. The dependent variable is the on-the-job test score achieved in nine different assessments, standardized to be mean zero and unitary standard deviation in the full sample of male workers. The score goes from 0 to 100 in each test, and each test is weighted by the credits assigned to it by the organization. The omitted category is the treatment group which received the female photograph and information on low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high returns to ability (specification (2) of Section 5.1). Columns (3) and (4) additionally control for an index of "difficulty" of the community where the worker is allocated to. For each local authority, I compute an index of "difficulty" by averaging the score in these variables: social workers' caseload, turnover, absenteeism and scores on helping children, child care, leadership effectiveness. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration, non-white ethnicity, workplace region and a dummy for being allocated to the preferred region. Columns (2) and (4) control for the index of observable qualifications which are positively correlated with receiving a job offer. Standard errors are clustered at the worker level.

Table 4. New male hires: attitudes on the job

DV:	(1) Perceived workload	(2) Concerned	(3) Satisfied	(4) Perceived at work	(5) impact outside	(6) Confidence in practice	(7) Intent to stay
Male Photo	0.010	0.097	0.352***	0.121	0.109	0.143	0.103
	(0.132)	(0.179)	(0.121)	(0.100)	(0.126)	(0.112)	(0.079)
High Exp Returns	0.016	-0.044	0.198**	0.368***	0.125	0.295**	0.267**
	(0.119)	(0.172)	(0.091)	(0.100)	(0.119)	(0.113)	(0.107)
Observations	83	89	89	67	66	89	89
R-squared	0.055	0.039	0.214	0.283	0.097	0.226	0.194
Basic Controls	✓	✓	✓	✓	✓	✓	✓
Survey Wave FE	✓	✓	✓	✓	✓	✓	✓
Mean Dep Var	0.31	0.50	0.36	0.55	0.09	0.43	0.57
Mean Dep Var in Pure C	0.44	0.59	0.63	0.67	0.25	0.67	0.89
Photo = Exp Ret p-val	0.97	0.58	0.33	0.05	0.94	0.32	0.26

Clustered s.e. in parentheses (worker level)
*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates for men only. The table reports results of seven different regressions. The omitted category is the treatment group which received the female photograph and information on low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high returns to ability (specification (2) of Section 5.1). "Perceived workload" is an indicator variable equal to one if a worker perceives their workload to be too high or much too high (Column 1). "Concerned" is an indicator variable equal to one if a worker feels any personal, financial, work, well-being or health-related concern (Column 2). "Satisfied" is a dummy for whether the person feels satisfied with their work (Column 3). "Perceived impact" is an indicator equal to one if a worker feels that he is having positive social impact in his work (Column 4) or outside work (Column 5). "Confidence in practice" is equal to one if the worker feels confident in interacting with families in need (Column 6). "Intent to stay" is an indicator equal to one if the worker says he is moderately or very likely to stay in the same community or in the same job (Column 7). All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration, non-white ethnicity and an additional dummy for the survey wave.

Table 5. Heterogeneous treatment effects by background and outside option

	DV: Applied and never DO = 1								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Heterogeneity by background				*Panel B: Heterogeneity by outside option*				
	Job Genderization		Men in Pink-Collar		Wage dispersion		Quantiles of outside option		
	High	Low	Low	High	Low	High	1^{st}	2^{nd}	3^{rd}
Male Photo	-0.026	-0.011	-0.010	-0.020	-0.014	-0.025	-0.060	-0.011	0.048
	(0.050)	(0.050)	(0.050)	(0.050)	(0.046)	(0.055)	(0.062)	(0.061)	(0.059)
High Exp Returns	0.167***	-0.021	0.112**	0.036	0.036	0.122**	0.103*	0.050	0.069
	(0.050)	(0.050)	(0.050)	(0.050)	(0.046)	(0.055)	(0.061)	(0.062)	(0.059)
Observations	390	402	394	398	477	330	260	266	281
R-squared	0.038	0.017	0.022	0.030	0.012	0.033	0.029	0.029	0.061
Basic Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean Dep Var	0.48	0.57	0.53	0.52	0.56	0.45	0.55	0.49	0.52
Photo = Exp Ret p-val	0.01	0.88	0.08	0.43	0.44	0.06	0.06	0.49	0.79
Rand Inf p-val									
Photo	0.58	0.80	0.86	0.70	0.74	0.67	0.34	0.84	0.41
Exp Returns	0.00	0.66	0.03	0.47	0.40	0.026	0.09	0.40	0.25

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates for men only. The omitted category is the treatment group which received the female photograph and information on low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high returns to ability (specification (2) of Section 5.1). All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration and non-white ethnicity. The rows "Rand Inf p-val" contain the p-values of the coefficients on the treatment dummies from randomization inference (randomization-t) with 1000 repetitions.

Panel A reports results of four different regressions. The variable "Job Genderization" is the Duncan index of occupational segregation by gender computed at the local area level (MSOA) where the subject went to secondary school or permanently lives. The variable "Men in Pink-Collar" is the average share of men in female-dominated jobs at the same local area level (MSOA). Both indexes are computed using data from the 2011 U.K. Census. I define "female-dominated occupations" the ones that have more than 75% of female workers for England overall. At the local level, I then computed the following average male proportion in those occupations as: $\sum_{i=1}^{N} \frac{m_i}{m_i+f_i}$, where m_i and f_i are respectively the number of men and women in female-dominated occupation i in a certain MSOA. For both indexes, the level "high" or "low" is defined for values of the index respectively above or below the gender-specific median in the experimental sample.

Panel B reports results of five different regressions. In Columns (5) and (6), wage dispersion for individual studying subject s is defined as the weighted average of the 75/25 interquartile range of the distribution of hourly wages across industries in the UK labour market, with weights are given by the proportion of graduates of subject s working in each industry.The level "high" or "low" is defined for values of the index respectively above or below the gender-specific median in the experimental sample. The outside option in Columns (7) to (9) is computed as the imputed expected wage in the UK labour market conditional on subject studied, gender, race, age, British nationality and marital status. Data are from the 2017 and 2018 UK Labour Force Survey.

Table 6. Women's results

	(1)	(2)	(3)	(4)	(5)	(6)
	Application and selection process			On-the-job outcomes		
VARIABLES	Applied and never DO	Received Offer	Accepted Offer	Mean on-the-job score Semester 1	Semesters 2/3	Early Exit
Male Photo	-0.051***	0.013	0.131**	1.570	-3.064	0.076*
	(0.017)	(0.015)	(0.055)	(1.149)	(2.058)	(0.042)
High Exp Returns	-0.015	0.004	-0.002	-0.465	-2.430	0.095**
	(0.017)	(0.015)	(0.055)	(1.124)	(2.114)	(0.043)
Observations	3,513	2,062	301	191	191	191
R-squared	0.013	0.025	0.028	0.020	0.026	0.056
Basic Controls	✓	✓	✓	✓	✓	✓
Mean Dep Var	0.60	0.14	0.55	57.3	60.1	0.02
Mean Dep Var in Pure C	0.59	0.15	0.68	59.5	57.8	0.07
Photo = Exp Ret p-val	0.12	0.67	0.08	0.21	0.81	0.74
Rand Inf p-val						
Photo	0.00	0.43	0.03	0.15	0.17	0.07
Exp Returns	0.36	0.83	0.96	0.88	0.19	0.03

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates for women only. The table reports results of six different regressions. The omitted category is the treatment group which received the female photograph and information on low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high returns to ability (specification (2) of Section 5.1). The dependent variables are indicator dummies for application, receiving a job offer (conditional on applying) and accepting the job offer (conditional on receiving the offer) in Columns (1), (2) and (3). The dependent variable in column (4) is the average on-the-job test score achieved in the first five assessments during the first semester on the job and in column (5) is the average on-the-job test score achieved in the four additional assessments during the second and third semester on the job. The score goes from 0 to 100 in each test, the average is weighted by the credits assigned to each exam by the organization. The dependent variable in column (6) is a dummy equal to one if the person left the programme before completing it. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration, non-white ethnicity. The rows "Rand Inf p-val" contain the p-values of the coefficients on the indicated treatment dummies from randomization inference (randomization-t) with 1000 repetitions.

Table 7. Overall gender shares and performance

	Male Share			**Share of Top Performers**		
	Applicants	Offerees	Workers	Women	Men	Overall
			Photograph			
Female Photo	18%	15%	16%	54%	67%	56%
Male Photo	17%	21%	21%	57%	42%	54%
			Information			
Low Exp Returns	16%	14%	14%	54%	40%	52%
High Exp Returns	19%	22%	**23%**	57%	58%	**57%**

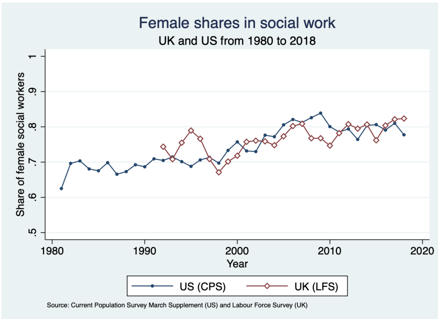
Note. The first three columns of this table show the male share among applicants (Column 1), people who received a job offer (Column 2) and workers (Column 3), excluding people who drop out before completing the programme. The last three columns show the share of top performers (people with an average score above 60%) by women (Column 4), men (Column 5) and for the pooled sample, excluding people who do not complete the programme.

Appendices

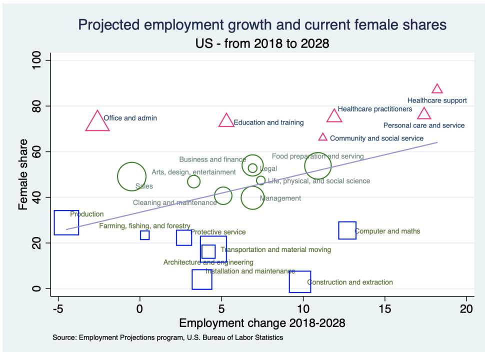
A Appendix figures and tables

Figure A.1. Three facts about social work

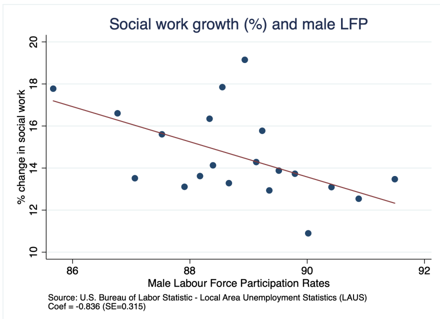
(a) Historical female shares in social work



(b) Predicted growth in occupations by current female shares (2018-2028)



(c) Social work predicted growth and male labour force participation



Note. Figure (a) shows the female share of social workers from 1980 to 2018 in the US (CPS data, March Supplement) and UK (LFS data). Figure (b) shows the correlation between predicted percentage change in employment between 2018 and 2028 for major occupational groups in the US (on the x-axis) and the 2018 female share (on the y-axis). Data are from the US Bureau of Labor Statistic and the Employment Projections program. Figure (c) shows a binned scatterplot between the 2018 male labour force participation (on the x-axis) and percentage change in employment in social work between 2018 and 2028 (on the y-axis) across US states. The graph controls for the overall growth rate across occupations and the state-level female labour force participation. Data are from the US Bureau of Labor Statistic Local Area Unemployment Statistics (LAUS) and the Employment Projections program.

Table A.1. Treatment effects: photographs and information interacted

	DV: Applied and never DO = 1	
	(1)	(2)
	Men	Women
(W,H)	0.088*	0.025
	(0.050)	(0.023)
(M,H)	0.066	-0.067***
	(0.049)	(0.024)
(W,L)	0.011	
	(0.050)	
(M,L)		-0.011
		(0.023)
Observations	807	3,513
R-squared	0.018	0.014
Basic Controls	✓	✓
Mean Dep Var	0.50	0.60
Tests of coefficient equality		
$(-g, H) = (g, H)$	0.65	0
$(-g, L) = (-g, H)$	0.12	0.02
$(W, L) = (M, H)$	0.27	0.12
Rand Inf p-val		
$(-g, H)$	0.08	0.01
(g, H)	0.15	0.27
$(-g, L)$	0.83	0.67

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates run separately for men (Column 1) and women (Column 2). For each gender g, the omitted category is the treatment group (g, L). Each regressor (P,S) is a treatment dummy for the combination of a male (M) or female (W) picture and high (H) or low (L) expected returns information. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration and non-white ethnicity. The rows "Rand Inf p-val" contain the p-values of the coefficients on the indicated treatment dummies from randomization inference (randomization-t) with 1000 repetitions.

Table A.2. Do women and men react differently to treatments?

VARIABLES	(1) Applied and never DO	(2) Received Offer	(3) Accepted Offer
Male Candidate	-0.103***	-0.045	0.056
	(0.033)	(0.029)	(0.128)
Male Photo	-0.051***	0.013	0.132**
	(0.017)	(0.015)	(0.055)
Male Photo x Male Candidate	0.034	0.040	-0.048
	(0.039)	(0.037)	(0.133)
High Exp Returns	-0.015	0.004	-0.002
	(0.017)	(0.015)	(0.055)
High Exp Returns x Male Candidate	0.087**	0.058	-0.055
	(0.039)	(0.037)	(0.132)
Observations	4,320	2,502	368
R-squared	0.015	0.029	0.025
Basic Controls	✓	✓	✓
Mean Dep Var	0.60	0.14	0.55

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates for the pooled sample of men and women. The omitted category is the treatment group which received the female photograph and information on low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high returns to ability (specification (2) of Section 5.1). The dependent variables are indicator dummies for application, receiving a job offer (conditional on applying) and accepting the job offer (conditional on receiving the offer) in Columns (1), (2) and (3). All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration, non-white ethnicity.

Table A.3. Treatment effects by outside option parameters (women)

	(1)	(2)	(3)	(4)	(5)	(6)
	Wage dispersion		Quantiles of outside option			
	Low	High	1st	2nd	3rd	4th
Male Photo	-0.063***	-0.017	-0.059*	-0.010	-0.065*	-0.068**
	(0.019)	(0.033)	(0.031)	(0.034)	(0.034)	(0.034)
High Exp Returns	-0.018	-0.009	-0.029	-0.037	-0.014	0.012
	(0.019)	(0.033)	(0.031)	(0.034)	(0.033)	(0.034)
Observations	2,619	894	937	828	874	874
R-squared	0.014	0.011	0.016	0.016	0.014	0.019
Basic controls	✓	✓	✓	✓	✓	✓
Mean Dep Var	0.62	0.56	0.71	0.58	0.60	0.51
Photo = Exp Ret p-val	0.10	0.86	0.50	0.56	0.29	0.09
Rand Inf p-val						
Male Photo	0.00	0.60	0.05	0.76	0.04	0.04
Exp Returns	0.34	0.77	0.31	0.29	0.663	0.73

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates for women only. The table reports results of five different regressions. The omitted category is the treatment group which received the female photograph and information on low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high returns to ability (specification (2) of Section 5.1). In Columns (1) and (2) wage dispersion is defined in the following way. For a candidate who studied subject s, the variable "Wage Dispersion" is computed as the weighted average of the 75/25 interquartile range of the distribution of hourly wages across industries in the UK labour market, where weights are given by the proportion of graduates of subject s working in each industry. The level "high" or "low" is defined for values of the index respectively above or below the gender-specific median in the experimental sample. The outside option in Columns (3) to (5) is is computed as the imputed expected wage in the UK labour market conditional on subject studied, gender, race, age, British nationality and marital status. Data are from the 2017 and 2018 UK Labour Force Survey. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration and non-white ethnicity. The rows "Rand Inf p-val" contain the p-values of the coefficients on the treatment dummies from randomization inference (randomization-t) with 1000 repetitions.

Table A.4. Employer's hiring criteria

	Information		Photographs	
	(1)		(2)	
DV:	Offer	p-val	Offer	p-val
Top University * T^1	0.054*	0.67	0.070**	0.69
	(0.028)		(0.028)	
Top University * T^2	0.071**		0.054*	
	(0.029)		(0.029)	
First Grade * T^1	0.110***	0.19	0.063**	0.04
	(0.032)		(0.030)	
First Grade * T^2	0.109***		0.160***	
	(0.032)		(0.034)	
Aligned Subject * T^1	-0.010	0.06	0.007	0.75
	(0.019)		(0.019)	
Aligned Subject * T^2	0.029		0.013	
	(0.020)		(0.020)	
Past Volunteering * T^1	0.047**	0.76	0.053***	0.85
	(0.020)		(0.020)	
Past Volunteering * T^2	0.056***		0.048**	
	(0.020)		(0.020)	
Maths Pre-Uni Score * T^1	0.004	0.31	-0.029	0.38
	(0.027)		(0.024)	
Maths Pre-Uni Score * T^2	-0.033		0.004	
	(0.026)		(0.029)	
English Pre-Uni Score * T^1	0.084***	0.57	0.089***	0.34
	(0.025)		(0.024)	
English Pre-Uni Score * T^2	0.064**		0.056**	
	(0.025)		(0.026)	
Observations	2,295		2,295	
R-squared	0.058		0.059	
Stratification Controls	✓		✓	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: OLS estimates. In Column (1), T^2 indicates information on high returns to ability (and T^1 the alternative information). In Column (2), T^2 indicates a male photograph (and T^1 a female photograph). All regressions include controls for gender and ethnicity (stratification variables). Independent variables are interacted with the treatment and control dummies. "Top University" is equal to one if the candidate attended a top tier university in the U.K. "First Grade" is equal to one if the candidate got a first grade in university. "Past Volunteering" is equal to one if the candidate volunteered frequently in the past."Maths Pre-Uni Score" and "English Pre-Uni Score" are equal to one if the candidate took the highest grade in Maths and English pre-university qualifications. The same results hold adding interactions for high cognitive and high manual skills, defined using the employment history reported by candidates in their application form. I find no differences in the extent to which the employer considers these skills desirable between treatments (p-vals > 0.14 for cognitive skills and p-vals > 0.4 for manual skills).

Table A.5. Effort in application completion

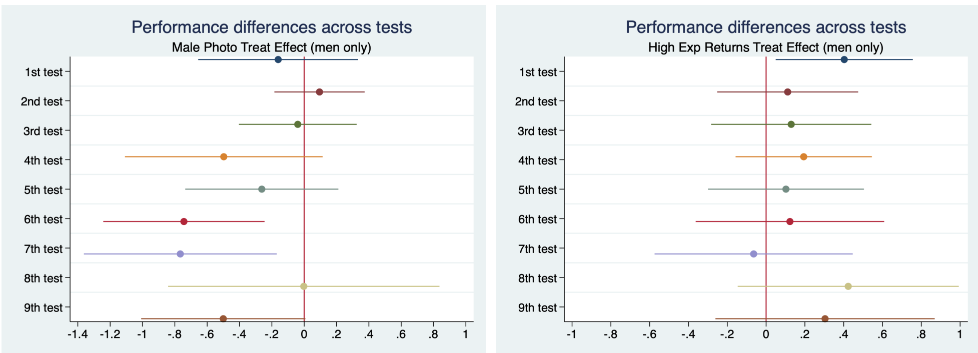
VARIABLES	(1) Access to portal	(2) # edits	(3) % completed	(4) Qst 1 length	(5) Qst 2 length
Male Photo	0.028	1.496	-0.022	-24.871	-36.319
	(0.025)	(2.110)	(0.023)	(55.613)	(46.020)
High Exp Returns	0.009	4.581**	0.027	34.469	42.934
	(0.025)	(2.205)	(0.023)	(55.683)	(46.085)
Observations	804	687	807	807	807
R-squared	0.034	0.043	0.031	0.023	0.028
Basic Controls	✓	✓	✓	✓	✓
Week dummies	✓	✓	✓	✓	✓

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates for men only. The omitted category is the treatment group that received information on low expected returns to ability. The variable "Access to portal" is a dummy for whether the person ever accessed the application portal to make changes to the application. The variable "# edits" counts how many times a candidate logged-in to make changes to the application form before submitting it. "% completed" is percentage of fields filled-in (not blank) in the application form. The variables "Qst 1 length" and "Qst 2 length" count number of characters used in each of the two motivational questions contained in the application form. All the regressions contain dummies for the week in which the candidate registered. The regressor "High Exp Returns" is a dummy equal to one for information on high expected returns to ability (specification (2) of Section 5.1). All the regressions control for the basic set of controls X_i: past application, access to early registration and non-white ethnicity.

Figure A.2. On-the-job test scores differences by treatment over time



Note. The figure reports the coefficients from a regression of each of the nine on-the-job assessment scores on the treatment dummy for receiving a male photograph (on the left) and the high expected returns statistics (on the right) for men only. Scores have been standardized by subtracting the gender-specific mean and dividing by the standard deviation. Coefficients are reported in chronological order from the top (first assessment) to the bottom (most recent assessment). All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration, non-white ethnicity, workplace region and allocation to preferred region.

Table A.6. Women's on-the-job performance

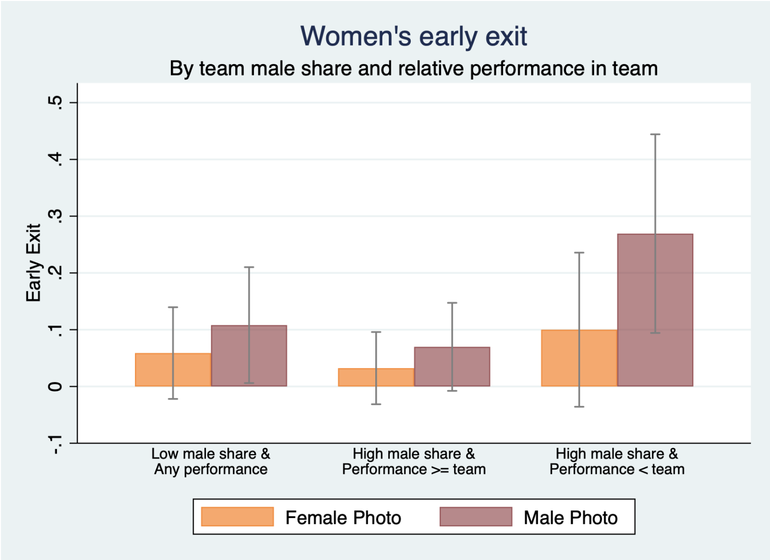
	DV: On-the-Job Std. Scores			
	(1)	(2)	(3)	(4)
Male Photo	-0.006	-0.024	-0.009	-0.025
	(0.094)	(0.087)	(0.096)	(0.089)
High Exp Returns	-0.118	-0.081	-0.117	-0.081
	(0.087)	(0.078)	(0.086)	(0.078)
Observations	1,716	1,716	1,716	1,716
R-squared	0.071	0.123	0.071	0.123
Basic Controls	✓	✓	✓	✓
Exam FE	✓	✓	✓	✓
Quality Controls	×	✓	×	✓
Location Difficulty Controls	×	×	✓	✓
Mean Dep Var	0.03	0.03	0.03	0.03
Mean Dep Var in Pure C	0.03	0.03	0.03	0.03
Photo = Exp Ret p-val	0.32	0.59	0.33	0.33

Clustered s.e. in parentheses (worker level)

*** p<0.01, ** p<0.05, * p<0.1

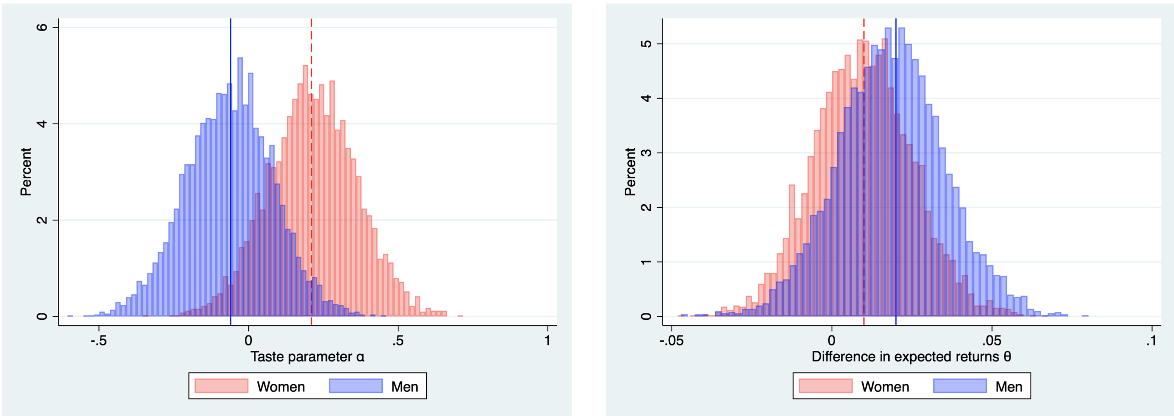
Note. OLS panel estimates for women only. The table reports results of four different regressions. The dependent variable is the on-the-job test score achieved in nine different assessments, standardized to be mean zero and unitary standard deviation in the full sample of male workers. The score goes from 0 to 100 in each test, and each test is weighted by the credits assigned to it by the organization. The omitted category is the treatment group which received the female photograph and information on low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high returns to ability (specification (2) of Section 5.1). Columns (3) and (4) additionally control for an index of "difficulty" of the community where the worker is allocated to. For each local authority, I compute an index of "difficulty" by averaging the score in these variables: social workers' caseload, turnover, absenteeism and scores on helping children, child care, leadership effectiveness. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration, non-white ethnicity, workplace region and a dummy for being allocated to the preferred region. Columns (2) and (4) control for the index of observable qualifications which are positively correlated with receiving a job offer. Standard errors are clustered at the worker level.

Figure A.3. Women's early exit from job by team characteristics



Note. The Figure shows the average rate of turnover among women, splitting the sample in three categories: i) women allocated in teams with median or below median male share ($\leq$ 20%) and any relative performance, ii) women allocated in teams with higher than median male share ($>$ 20%) and with individual performance which is better than the leave-one-out team average and iii) women allocated in teams with higher than median male share ($>$ 20%) and with individual performance which is worse than the leave-one-out team average.

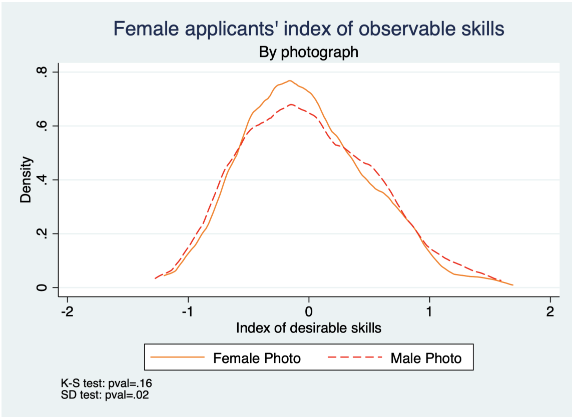
Figure A.4. Estimated weight on workplace gender shares and change in expected returns to ability



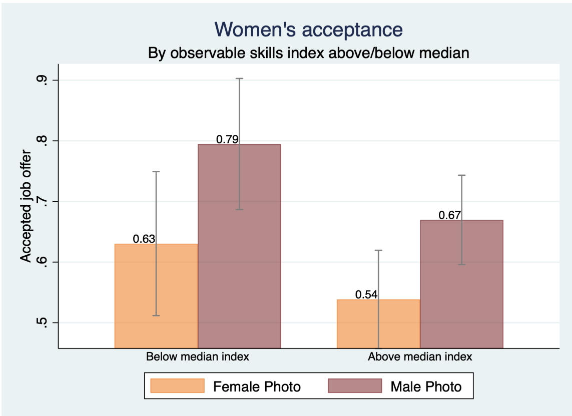
Note. The figure on the left-hand side shows distributions of the estimated weight on workplace gender shares α. The figure on the right-hand side shows distributions of the estimated difference in expected returns to ability $\Delta\theta$. Both graphs use the discrete-choice framework of section 7.3. Darker (blue) bars are for men and lighter (red) bars are for women. Solid lines are the mean value of the parameters for men and dashed lines are the mean value of the parameters for women. Multiple estimations are obtained through 5000 bootstrap replications of the logit model described in the main body of the paper with equal sample size (N=800) for the two genders.

Figure A.5. Female applicants' skills and offer acceptance

(a) Female applicants' skills

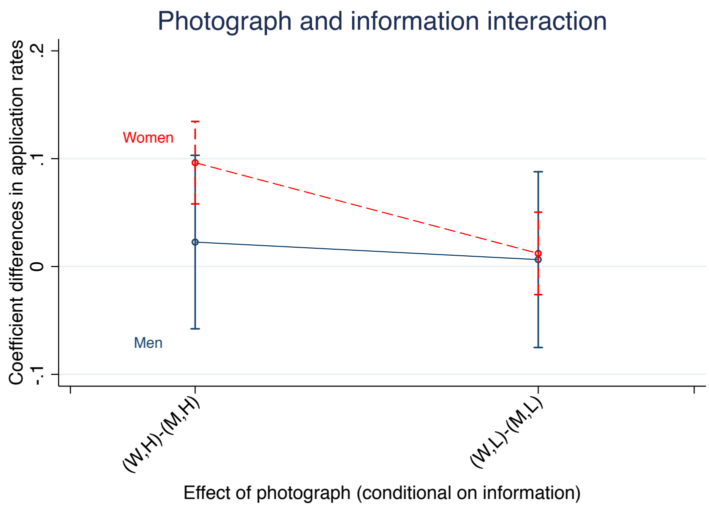


(b) Women's offer acceptance by skills



Note. The figure shows the distribution of female applicants' observable skills (a) and female applicants' offer acceptance by their level of observable skills (b). I build an index of observable skills (”desirable-skillset index”) computed as the mean of the following standardized variables: receiving a first grade, being from a top tier university, frequent past volunteering, high cognitive skills and score in English pre-university tests. The figure in panel (a) shows the distribution of this index by photograph treatment among women who apply for the position. The figure in panel (b) shows the acceptance rate for women by above/below median “desirable skills index” and by photograph treatment, conditional on receiving a job offer.

Figure A.6. Interaction between photographs and information on applications



Note. The figure shows the estimated difference in application rates between the male and female photograph treatments conditional on each type of information. That is, $(W, s) - (M, s)$ with $s \in \{H, L\}$. Dashed (red) lines are for women and solid (blue) lines are for men.

Table A.7. Attention to experimental emails

DV: Never asked for reminder				
	(1)	(2)	(3)	(4)
	Men		Women	
Male Photo	-0.071**	-0.069**	0.004	0.004
	(0.028)	(0.028)	(0.013)	(0.013)
High Exp Returns	-0.042	-0.040	-0.030**	-0.030**
	(0.028)	(0.028)	(0.013)	(0.013)
Observations	799	799	3,476	3,476
R-squared	0.038	0.042	0.023	0.024
Basic Controls	✓	✓	✓	✓
Outside Option Control	×	✓	×	✓
Mean Dep Var	0.84	0.84	0.80	0.80

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates. The dependent variables is a dummy equal to one if the candidate never asked for a reminder of his/her unique candidate number, which is needed to access the application portal and is shown in the invitation-to-apply email. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information on high expected returns to ability (specification (2) of Section 5.1). All the regressions control for the basic set of controls X_i (past application, access to early registration, non-white ethnicity) and for the number of times the candidate accessed the application portal.

Table A.8. Treatment effects by sexuality and marital status

DV: Applied and never DO				
	(1)	(2)	(3)	(4)
	Women		Men	
Male Photo	-0.065***	-0.054***	-0.055	-0.037
	(0.017)	(0.017)	(0.038)	(0.039)
Non Hetero	0.010		-0.130*	
	(0.049)		(0.070)	
Male Photo * Non Hetero	0.080		0.148	
	(0.064)		(0.105)	
Married		0.002		-0.020
		(0.035)		(0.066)
Male Photo * Married		-0.011		0.018
		(0.051)		(0.088)
Observations	3,294	3,455	757	793
R-squared	0.022	0.021	0.022	0.020
Basic controls	Y	Y	Y	Y
Mean Dep Var	0.60	0.61	0.53	0.54

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

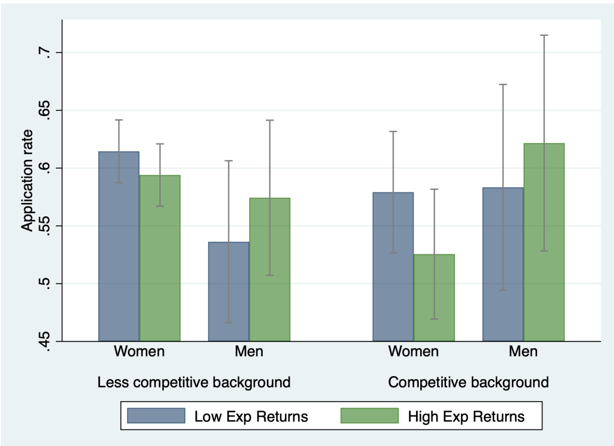
Note. OLS estimates. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment. "Non hetero" is a dummy equal to one if the person stated to be non-heterosexual and missing for refusing to answer the question on sexuality. "Married" is a dummy for being married or in a civil partnership and missing for refusing to answer the question on marital status. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration and non-white ethnicity.

Table A.9. A measure of overconfidence by gender

Overconfidence: self-reported number of skills above the mean							
	Women			Men			
	Mean	SD	N	Mean	SD	N	p-val
			Overall				
General	5.63	2.84	548	5.36	2.96	85	.43
Job specific	2.92	1.63	548	2.49	1.7	85	.03**
			Pure Control only				
General	5.5	2.73	123	5.63	2.95	19	.85
Job specific	2.82	1.55	123	2.53	1.84	19	.45

Note. The measure of overconfidence is defined in the following way. I asked survey respondents (N=633) to rate themselves in ten skills on a scale from 1 (min) to 10 (max). The skills are both general (i.e. complex problem solving, finance management, critical thinking, creativity, adaptability) and job specific (active listening, effective communication, leadership, empathy, client support). For each person, I construct a measure of overconfidence by counting the number of skills rated above the sample mean. The Table shows the mean measure of overconfidence by gender across treatments (in the first two rows) and in the pure control only (last two rows). Survey respondents are the subset of field participants who responded to a survey invitation (11.4%) sent to everyone in the invitation-to-apply email and subsequently encouraged through an ad-hoc email adding monetary incentives. The survey sample is representative of the overall pool of candidates (e.g., balanced on gender, treatment assignment, FTE status).

Figure A.7. Shock to expectations and competitiveness



Note. The graph shows raw differences in application rates in the high and low expected returns treatments by gender and a proxy of competitive attitudes. The proxy of competitive attitudes is built using information on the candidates' educational background. "Competitive background" is defined as having studied a male-dominated subject (e.g., engineering, business, math) in a top tier university in the UK. "Less competitive background" is defined as having studied a female-dominated subject (e.g., psychology, languages, humanities) in a non top tier university. "Female" and "Male" indicate the candidates' gender.

B Auxiliary online experiments

In this section I first address treatment-specific issues which relate to differences in pictures' content and the interpretation of the information provided. I use auxiliary survey evidence that I have collected on three different samples between July and December 2018. I then turn my attention to issues that might affect results equally across treatments.

B.1 Treatment-specific threats

The main goal of this section is to check for differences between photographs (messages) used in the intervention which might confound the interpretation of the results. For instance, photographs might not differ only in the subjects' gender, but also in their expression, clothes and other observable or unobservable characteristics. Regarding information, one might worry that the sentences reporting statistics of past performance could be interpreted as signals of other job amenities (e.g., wage).

Sampling

In July 2018, I conducted checks on differences between photographs on a sample of 161 Amazon Turk workers. This allows me to understand whether images differed in some important dimensions other than gender, but correlated with it. Between November and December 2018 I administered an online survey to 565 people in the UK to understand whether - and how - the intervention emails affect their beliefs about the job and its applicants. In a between-subject design, I first showed respondents a photograph and asked two short questions about the portrayed worker (from the previous survey on Amazon Turk). Then participants looked at one intervention email for some time (at least 30 seconds).[72] After mandatory understanding checks, I elicited beliefs on a variety of dimensions about the job and its applicants (e.g., wage, difficulty). I implemented the survey using two samples of respondents: 2018/2019 applicants of the partner organization and workers on the platform "Prolific Academic". The sampling strategy maximizes the similarity to my field sample.The out-of-trial sample of applicants for the same organization is meant to capture unobservable characteristics that people interested in the particular job and/or organization share. However, the number of male respondents is too small to allow analyses by gender. I selected the sample on Prolific Academic by matching the composition of the field sample on several observables criteria. Participation was incentivized and average completion time was 15 minutes. The following paragraphs describe the sampling in detail.[73]

Amazon Turk photographs categorization. Respondents were Amazon Mechanical Turk workers who had not participated in any of the researchers' previous experiments conducted on the same platform and who had been granted the "Master" qualification on the website. The survey was conducted with a pool of workers all around the world. The survey was run in different waves between May and July 2018. A total of 188 answers were collected (on average 47 per photograph) and I excluded answers which were only partial (with less than 95% completed). The final sample is made up of 161 answers, of which 39 were for the white woman, 38 for the white man and 42 for the non white photographs. The survey took an average of 2 minutes and was rewarded with 20 cents.

[72] The intervention table was shown, as in Figure 3.

[73] I registered pre-analysis plans before conducting analyses on these survey data.

2018 Applicants sample. At the beginning of November 2018, I collaborated with the partner organization to invite current candidates to participate in my online survey. Invitations were sent to 4500 people over two days. The sample comprises candidates at different stages of the selection process who registered between the beginning of September and the beginning of November.[74] As incentive for participation I compensated the first 300 respondents with £5, which they could keep for themselves or donate to a UK social work charity[75] All the participants were also automatically enrolled into a raffle for a £150 Amazon voucher. A total of 303 people fully completed the survey, which corresponds to a response rate of around 7%. While men's proportion corresponds to the population mean - less than 20% - their number is too small to allow analyses by gender in this sample.

Prolific Academic sample. Respondents in this sample are Prolific Academic workers who i) have not participated in any of the researchers' previous surveys conducted on the same platform, ii) are of British nationality, iii) have an approval rate between 75 and 100 percent, iv) are between 18 and 64 years old and v) have at least a bachelor degree. The final sample is made up of 130 women and 131 men, selected through independent survey postings on the website. I collected answers in different waves to match the composition of the field sample on the following observable criteria: gender, ethnicity, student status, university subject, employment status, job sector. Payment was £1.50.

Photographs checks

In the Amazon Turk photographs categorization task, I asked respondents to rate photographs along the following dimensions: friendliness, work satisfaction, emotions evoked, trustworthiness, attractiveness and clothing. In the other two samples, I asked respondents to categorize the people portrayed in the intervention photographs along two characteristics: friendliness and work satisfaction. Each respondent was asked about only one photograph, which was the same used afterwards in displaying the full intervention. Table B.1 presents mean differences between the male and female photographs within each pair of white and non-white photographs. The table below shows that women and men's pictures were rated similarly in most dimensions, but there is a significant and consistent difference in terms of perceived friendliness in the photos portraying white people. Such a difference, however, cannot explain the results, which are the same for both white and non-white candidates.

Information checks

In addition to the manipulation checks reported in the main body of the paper, I elicited respondents' beliefs about success on the job by asking the following question: "After seeing the email ad, please indicate below the proportion of [women/men] that you think are successful on-the-job. Interpret "success" as people who got commendable or excellent feedback on the job." I construct a variable for the average percentage of high-performers on the job by weighting the answers to the gender-specific questions (with 0.8 and 0.2 weights for women and men respectively). I similarly construct a variable for the beliefs about the quality of the pool of applicants with the following question: "Consider 100 [women/men] that apply for this job in social work after seeing the email ad. How many do you

[74] The sample includes registered candidates who have yet to submit the application form, applicants who passed the first stage of the selection process and candidates already rejected.

[75] Participants could select one out of two social work charities for the donation.

Table B.1. Photographs: manipulation checks

	Female Photo			Male Photo			Diff means
	Mean	SD	N	Mean	SD	N	P-val
Panel A: 2018 Applicants							
White pictures							
Friendliness	.79	.41	92	.63	.48	95	.01
Work satisfaction	.91	.28	92	.84	.37	95	.14
Non-white pictures							
Friendliness	.86	.36	28	.82	.39	28	.72
Work satisfaction	.82	.39	28	.93	.26	28	.23
Panel B: Prolific Ac sample							
White pictures							
Friendliness	.87	.34	98	.74	.44	95	.02
Work satisfaction	.81	.4	98	.76	.43	95	.42
Non-white pictures							
Friendliness	.97	.17	33	.92	.28	36	.35
Work satisfaction	.97	.17	33	.92	.28	36	.35
Panel C: Amazon Turk sample							
White pictures							
Happy feeling	.79	.41	39	.66	.48	38	.18
Friendliness	.9	.31	39	.74	.45	38	.07
Work satisfaction	.87	.34	39	.76	.43	38	.22
Trust	.85	.37	39	.82	.39	38	.73
Attractiveness	.72	.46	39	.76	.43	38	.66
Professional clothing	.38	.49	39	.87	.34	38	0
Non-white pictures							
Happy feeling	.9	.3	42	.9	.3	42	1
Friendliness	.98	.15	42	.95	.22	42	.56
Work satisfaction	.95	.22	42	.88	.33	42	.24
Trust	.93	.26	42	.88	.33	42	.46
Attractiveness	.95	.22	42	.74	.45	42	.01
Professional clothing	.93	.26	42	.9	.3	42	.7

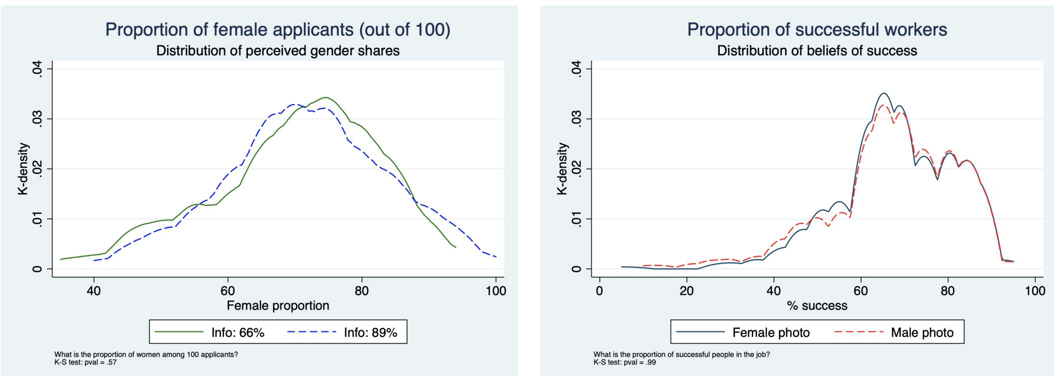
Note. Friendliness of the person in the picture was rated answering the question: "How does the person in the photograph appear to you?" on a 5-points scale. The variable "Friendliness" is a dummy equal to 1 if the person replied Friendly or Very Friendly and 0 otherwise. Work satisfaction was rated answering: "In your opinion, how satisfied is this person in his/her work?" on a 5-points scale. The variable "Work Satisfaction" is a dummy equal to 1 if the person replied Satisfied or Very Satisfied and 0 otherwise. The question "To what extent does this image make you feel happy?" assessed emotional reaction to the picture on a 7-points scale. The variable "Happy feeling" takes values between -3 ("Extremely unhappy") and 3 ("Extremely happy"). The variable for trust is defined from answers to the question "If this person was giving you some information about her job, would you trust him/her?", to which people answered on a 5-points scale; the variable has values between -2 ("Definitely not") and 2 ("Definitely yes"). The variable attractiveness is defined from answers to the question "In your opinion, how does this person look like?", to which people answered on a 5-points scale; the variable has values between -2 ("Not attractive") and 2 ("Attractive"). The variable professional clothing is a dummy equal to one if the respondent would describe the clothes of the portrayed person as "professional" and 0 if "unprofessional". In the Amazon Mechanical Turk sample the number of respondents for each question may vary by design: the more sensitive questions on clothing, ethnicity, attractiveness and trust were asked only on a subset of respondents.

think that have the potential to get commendable or excellent feedback on the job?". To check for possible confounders in the interpretation of the email content, I then ask respondents to rate the job on different dimensions on a scale from 1 to 100. For instance, I asked them: "By looking at this ad, do you think that the job has a high or low wage? Indicate your answer on a scale from 0 (low wage) to 100 (high wage)".

Table B.2 shows mean differences in ratings between the two information treatments on the following job characteristics: wage, difficulty of job tasks, difficulty of promotion, number of applicants (out of 100 interested people) and proportion of female applicants (out of 100 applicants). Table B.2 also shows mean differences in people's opinion on whether the job is desirable for a man, whether the job is desirable for a woman, whether they think that customers discriminate against workers (by race or gender) and whether the job has a high social status. The answer was given on a 6-points Likert scale: I code the variables in the tables as 1 if people answer that they strongly agree, agree or slightly agree with the statement and 0 otherwise.

Table B.2 shows that respondents' beliefs about the quality of the pool of applicants and percentage of high performers in the job change according to the experimental information treatment. The sample of current applicants also slightly updates on job difficulty, social status and discrimination by customers, but the magnitude of these differences are small. Table B.3 shows that pictures do not affect updating on job amenities or quality of the pool, except for desirability by gender and the female proportion of applicants. Overall, this evidence supports the interpretation of the treatments given in the paper. Figure B.1 further checks whether information on past performance affects perceived gender proportion (graph on the left) and whether photographs affect updating on the proportion of successful people in the job. This is to exclude the possibility that the two treatments are interacting, which would make the identification of the two separate channels difficult.

Figure B.1. Interaction between photographs and information: manipulation checks



Note. The left panel shows the distribution of answers to the question "Consider 100 people who apply for this job. How many do you think are women?", separately for respondents assigned to the email with a high or low information on returns to ability. The right panel shows the distribution of answers to the question "After seeing the email ad, please indicate below the proportion of [WOMEN/MEN] that you think are successful on-the-job", separately for respondents assigned to the email with a female or male photograph. Data are from the auxiliary online surveys. The number of respondents is 504: 262 are from the Prolific Academic sample and 242 from the organization's sample.

B.2 Threats across treatments

There are two main concerns: people's attention to the intervention and participants' trust in the information presented. First, I cannot exclude that some people did not open the invitation-to-apply email, but unfortunately I do not have metrics on opening rates. If the decision to not open the email is negatively correlated with interest in applying, then the compliers to my intervention would be people with a higher baseline interest in the job. However, the correlation could also go the opposite way: the invitation-to-apply email contains a detailed description of the selection process that the least informed people might be interested in.

Overall, not opening the email is very unlikely: the invitation-to-apply email contains the candidate's unique reference number, which is essential to be able to access the application portal, submit the application form and have access to other steps of the process. In the overall sample, 15% of men and 13% of women never accessed the application portal, which is the upper bound of the proportion of people that might have not opened the email. The randomization should guarantee that proportion of "types" who did not look at the invitation-to-apply email is equally likely across experimental conditions, which should then only create an attenuation bias in the results.[76]

Another risk is that people did not pay attention to the intervention. There are two main ways in which attention could affect the results. If attention is an individual trait, such that some people are more attentive than others, it should not introduce any bias as long as it is balanced across treatments. If attention is instead endogenously chosen by experimental subjects, it becomes an outcome of the treatment, which should be considered as a potential confounder (see Section 9.2).

The experiment was designed also to limit inattention. The intervention box was located in the top quarter of the email and could be visualized in the email preview in any smartphone or tablet. It was also positioned right below the candidate number, which is one of the most important pieces of information contained in the invitation-to-apply email. Finally, the text on the right of the picture addressed the candidates by name to visually capture their attention (see Figure 2).

Participants' lack of trust in the experimenters (i.e. the organization) can limit the experiment's validity. The invitation-to-apply email was signed by the Director of Selection, it contained the organization's logo and a disclaimer of confidentiality. Participants were told that they could contact any member of the recruitment team for questions, which in principle include doubts about the information presented in the treatment emails.[77] Qualitative interviews with candidates indicate that they had not been surprised by seeing an email containing statistics about on-the-job performance. The organization is indeed well-known for its efforts to be evidence-based and statistics are frequently reported on the organization website.

[76] I cannot test this directly as the decision to access the application portal is endogenous and could be an outcome of the intervention itself. However, I computed Lee bounds for the treatment effects (Lee, 2009) for the extreme case that attrition involves all the people who never accessed the portal. For men, bounds for the effect of high expected returns to ability are tight and the effect confidence interval doesn't cover zero. The lower and upper bound are respectively .073 and .082, both statistically significant (p-val < 0.05). For women, bounds for the effect of the male photograph are less tight and the effect confidence interval covers zero at the upper bound. The lower and upper bound are respectively -0.06 and -0.02, with only the lower bound statistically significant (p-val < 0.005).

[77] To the best of my knowledge, this never happened.

Table B.2. Information and inference on job amenities

	66% Info Mean	89% Info Mean	Diff H-L	66% Info N	89% Info N
Panel A: 2018 Applicants sample					
Job difficulty	65.81	60.31	-5.49**	120	121
	(17.69)	(21.25)	(2.52)		
Wage level	51.14	51.32	0.18	43	41
	(12.88)	(15.76)	(3.13)		
Promotion difficulty	55.46	55.98	0.52	120	120
	(15.77)	(18.04)	(2.19)		
Job desirable for men	0.71	0.74	0.03	120	121
	(0.46)	(0.44)	(0.06)		
Job desirable for women	0.81	0.88	0.07	120	121
	(0.40)	(0.33)	(0.05)		
Discrimination by customers	0.39	0.53	0.14**	120	121
	(0.49)	(0.50)	(0.06)		
Job high social status	0.51	0.68	0.17***	120	121
	(0.50)	(0.47)	(0.06)		
% of high-skilled applicants	72.63	80.27	7.64***	120	122
	(19.62)	(20.05)	(2.55)		
% of high-performers on the job	68.20	73.72	5.52***	120	122
	(11.95)	(14.08)	(1.68)		
Number of applicants	61.72	58.36	-3.36	120	122
	(17.74)	(19.26)	(2.38)		
% female applicants	69.17	70.49	1.32	120	122
	(13.60)	(12.73)	(1.69)		
Panel B: Prolific Ac sample					
Job difficulty	65.61	62.51	-3.10	130	132
	(19.82)	(19.56)	(2.43)		
Wage level	43.95	45.95	2.00	130	132
	(19.64)	(17.59)	(2.30)		
Promotion difficulty	54.29	56.20	1.91	130	132
	(16.30)	(17.77)	(2.11)		
Job desirable for men	0.69	0.61	-0.08	130	132
	(0.46)	(0.49)	(0.06)		
Job desirable for women	0.95	0.93	-0.01	130	132
	(0.23)	(0.25)	(0.03)		
Discrimination by customers	0.45	0.41	-0.04	130	132
	(0.50)	(0.49)	(0.06)		
Job high social status	0.46	0.49	0.03	130	132
	(0.50)	(0.50)	(0.06)		
% of high-skilled applicants	65.81	76.62	10.81***	130	132
	(16.43)	(17.99)	(2.13)		
% of high-performers on the job	64.02	74.03	10.01***	130	132
	(12.42)	(12.26)	(1.53)		
Number of applicants	47.85	51.58	3.73	130	132
	(21.83)	(22.49)	(2.74)		
% female applicants	71.32	72.87	1.56	130	132
	(11.09)	(11.62)	(1.40)		

Note. On a scale from 0 to 100, participants are asked to what extent they think that the job i) is difficult, ii) has a high wage, iii) people get easily promoted. Rows 4 to 7 report the extent to which respondents agreed with the following statements: "the job is desirable for a man", "customers discriminate workers (by race or gender) in this job", "the job is desirable for a woman", "the job has a high social status". Answers were on a 6-points scale from "Strongly Agree" to "Strongly Disagree" and I created dummy variables equal to one for the three highest options. The variable "% of high-performers in the job" is the weighted average of answers to the questions "Now that you have seen the email ad...indicate below the proportion of [women/men] that you think are successful on-the-job". The variable "% of high-skilled applicants" is the weighted average of answers to the questions "Out of 100 [women/men] that apply for this job after seeing the email ad, how many do you think that have the potential to get commendable or excellent feedback on the job?". "Number of applicants" is the believed number of people that apply out of 100 who are considering whether or not to apply for the job. "% female applicants" is the perceived female share among 100 applicants. Some questions were shown to subsamples only, implying differences in the number of respondents.

Table B.3. Photographs and inference on job amenities

	Female Ph. Mean	Male Ph. Mean	Diff M-W	Female Ph. N	Male Ph. N
Panel A: 2018 Applicants sample					
Job difficulty	63.09	63.01	-0.08	119	122
	(20.34)	(19.16)	(2.55)		
Wage level	52.09	50.27	-1.82	44	40
	(16.10)	(12.09)	(3.13)		
Promotion difficulty	56.52	54.93	-1.59	119	121
	(17.22)	(16.63)	(2.19)		
Job desirable for men	0.62	0.82	0.19***	120	121
	(0.49)	(0.39)	(0.06)		
Job desirable for women	0.96	0.73	-0.23***	120	121
	(0.20)	(0.45)	(0.04)		
Discrimination by customers	0.51	0.41	-0.10	120	121
	(0.50)	(0.49)	(0.06)		
Job high social status	0.63	0.55	-0.08	120	121
	(0.48)	(0.50)	(0.06)		
% of high-skilled applicants	77.35	75.63	-1.73	120	122
	(19.11)	(21.20)	(2.60)		
% of high-performers on the job	72.60	69.39	-3.21*	120	122
	(11.99)	(14.40)	(1.70)		
Number of applicants	60.61	59.45	-1.16	120	122
	(18.56)	(18.62)	(2.39)		
% female applicants	72.50	67.22	-5.28***	120	122
	(12.60)	(13.22)	(1.66)		
Panel B: Prolific Ac sample					
Job difficulty	65.13	62.96	-2.17	131	131
	(18.67)	(20.71)	(2.44)		
Wage level	44.28	45.63	1.34	131	131
	(19.73)	(17.51)	(2.30)		
Promotion difficulty	53.56	56.95	3.40	131	131
	(17.62)	(16.36)	(2.10)		
Job desirable for men	0.60	0.70	0.10*	131	131
	(0.49)	(0.46)	(0.06)		
Job desirable for women	0.94	0.94	0.00	131	131
	(0.24)	(0.24)	(0.03)		
Discrimination by customers	0.47	0.38	-0.09	131	131
	(0.50)	(0.49)	(0.06)		
Job high social status	0.45	0.50	0.05	131	131
	(0.50)	(0.50)	(0.06)		
% of high-skilled applicants	70.50	72.02	1.52	131	131
	(18.71)	(17.36)	(2.23)		
% of high-performers on the job	68.48	69.64	1.16	131	131
	(12.96)	(13.66)	(1.65)		
Number of applicants	49.76	49.70	-0.06	131	131
	(22.59)	(21.89)	(2.75)		
% female applicants	74.91	69.29	-5.62***	131	131
	(10.62)	(11.43)	(1.36)		

Note. On a scale from 0 to 100, participants are asked to what extent they think that the job i) is difficult, ii) has a high wage, iii) people get easily promoted. Rows 4 to 7 report the extent to which respondents agreed with the following statements: "the job is desirable for a man", "customers discriminate workers (by race or gender) in this job", "the job is desirable for a woman", "the job has a high social status". Answers were on a 6-points scale from "Strongly Agree" to "Strongly Disagree" and I created dummy variables equal to one for the three highest options. The variable "% of high-performers in the job" is the weighted average of answers to the questions "Now that you have seen the email ad...indicate below the proportion of [women/men] that you think are successful on-the-job". The variable "% of high-skilled applicants" is the weighted average of answers to the questions "Out of 100 [women/men] that apply for this job after seeing the email ad, how many do you think that have the potential to get commendable or excellent feedback on the job?". "Number of applicants" is the believed number of people that apply out of 100 who are considering whether or not to apply for the job. "% female applicants" is the perceived female share among 100 applicants. Some questions were shown to subsamples only, implying differences in the number of respondents.

C Exposure to occupational gender segregation

The validity of the proxies for α used in Section 7.1 relies on the positive correlation between labour market genderization, social norms regarding men and women's career choices and beliefs about their skills in different occupations. There is a well-known relationship between occupational gender segregation and the gender wage-gap (Blau et al., 2013). Moreover, sociologists have been extensively studying the association between the former measure and gender attitudes (England, 2017). I present three data exercises to validate the proxy used.

First, I show that men who come from areas with above-median occupational gender segregation display a higher implicit association between social work and women. In the invitation-to-apply email, all the experimental subjects were invited to participate in a complementary research survey, which included a Single-Target Implicit Association test (Greenwald et al., 1998).[78] I designed an ad-hoc test to measure the extent to which respondents automatically associate social work with women.[79]

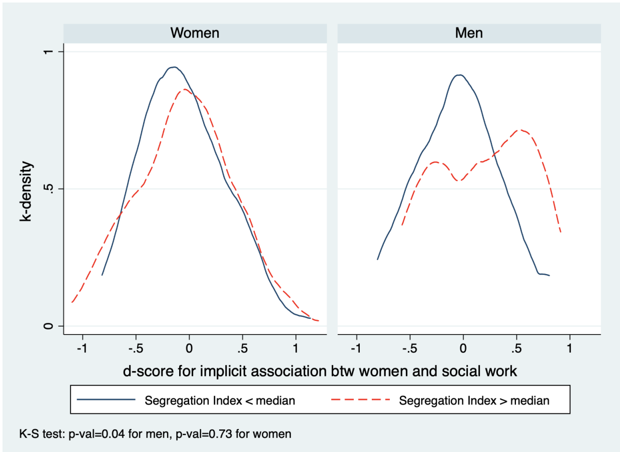
Subjects are presented with two sets of stimuli. The first set of stimuli are typical English female names (e.g., Rebecca) and male names (e.g., Josh), and the second set are words related to social work (e.g., family assistance). One word at a time appears on the screen and individuals are instructed to categorize it to the left or the right according to different labels displayed on the top of the screen (for instance, the respondent should categorize the word "Josh" either to the right - where the label is "Female" - or to the left - where the label is "Male"). Subjects are required to categorize the words as quickly as possible for four rounds. There are two types of rounds. In "hypothesis-inconsistent" rounds individuals categorize to one side of the screen female names and to the opposite side of the screen male names and social work activities. In "hypothesis-consistent" rounds individuals categorize to one side of the screen male names and to the opposite side of the screen female names and social work activities. The measure of implicit association between female gender and social work is given by the standardized mean difference score of the "hypothesis-inconsistent" rounds and "hypothesis-consistent" rounds. The intuition behind the test is that people with a greater implicit association of the job with women take longer to correctly categorize names in the "hypothesis-inconsistent pairings", because of the cognitive cost imposed by the inconsistent pairing of the two concepts. Thus, the higher and more positive the d-score the stronger is the association between the two concepts.[80]

Figure C.1 shows the distribution of d-score for women (left panel) and men (right panel), splitting the sample according to exposure to different levels of the Duncan Index. The distribution of d-score values for men exposed to higher-than-median gender segregation is strikingly shifted to the right of the distribution of men from lower-than-median gender segregation (Kolgorov-Smirnov test: p-val=0.043). A similar pattern is observed for women, but the difference is smaller and I cannot reject the null hypothesis of equal distribution between the groups (Kolgorov-Smirnov test: p-val=0.73). The null result of the photograph manipulation on men's applications is surprising in light of this evidence. A few recent economics papers show that implicit biases against minorities (by race or gender) are correlated with actual behaviour by managers (Glover et al., 2017) and teachers (Carlana, 2019). I provide evidence that labour market conditions correlate with implicit biases held by the minority,

[78] Response rate was 12.5% for the main survey and 6% to the IAT (604 and 300 respondents respectively).
[79] The IAT has been increasingly used in economics. For a review, see Bertrand and Duflo (2017).
[80] The order of the two types of blocks was randomized at the individual level.

but I do not find evidence for behavioural consequences.

Figure C.1. Implicit Association Test and exposure to gender occupational segregation



Note. The figure shows kernel density estimates of the d-score computed from an Implicit Association Test (IAT) I administered to the job candidates as part of a research survey (12% response rate). Respondents to the IAT count 337 women and 52 men (61% of the survey respondents). The d-score measures the degree of implicit association between female gender and social work: the higher and positive, the greater the implicit association. The d-score is the standardized mean difference score of the "hypothesis-inconsistent" rounds and "hypothesis-consistent" rounds. In the former type of rounds, individuals are instructed to categorize to one side of the screen female names and to the opposite side of the screen male names and social work activities ("hypothesis-inconsistent pairings"). The latter are rounds in which individuals must categorize to one side of the screen female names and social work activities and to the opposite side of the screen male names only ("hypothesis-consistent pairings").

In Figure C.2 I show that UK regions with high gender segregation levels display more traditional norms related to women's employment. In the two scatter plots of Figure C.2, the x-axis shows the proportion of local authorities in a certain region that have a Duncan Index in the top quartile of the national distribution. The y-axis shows the regional proportion of people who think that women are less successful than men in starting their own business (left panel) and that men should have priority in hiring when jobs are scarce (right panel). I use data from the 2013 British Attitudes Survey in the left figure and the 1995 and 2005 waves of the World Value Survey in the right figure.

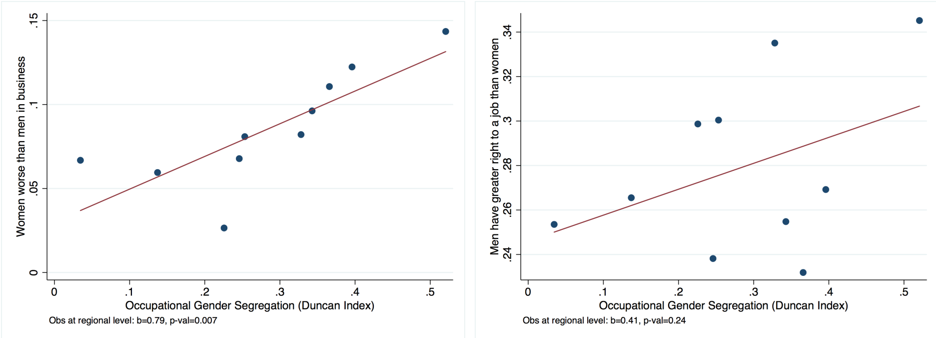
Table C.1 uses data from the auxiliary online experiments (described in Section B) to show whether people exposed to areas of high gender occupational segregation differ in terms of beliefs on men and women's skills in female occupations. In the surveys, I asked people the following questions:

a) On a scale from 0 (min) to 100 (max), what do you think is the performance of a [woman/man] in social work? (0 = extremely bad, 50 = neither bad nor good, 100 = extremely good)

b) On a scale from 0 (min) to 100 (max), how confident are you of your answer - that the performance of a [woman/man] in social work is [selection from a)]?

I use answers to the former question as a proxy for the priors on male and female performance in social work and to the latter as a proxy of priors' precision. The proxy for precision is the dependent variable in Table C.1, computed as an average of the precision levels stated in the two questions (one about a man and one about a woman). The independent variable is an indicator variable for a higher than median Duncan index of the postcode where a respondent was living when she/he was

Figure C.2. Correlation between gender occupational segregation and norms



Note. In both scatter plots, the variable on the x-axis is the proportion of census areas (MSOAs) within a region which have a value of the Duncan index above the 75^{th} percentile of the U.K. distribution. It is thus a measure of regional occupational gender segregation. Data are from the 2011 U.K. Census. In the left graph, the variable on the y-axis is the proportion of people in the region that replied "Slightly less successful" or "Much less successful" to the question: "Compared to men, how successful do you think women in general would be in setting up their own businesses?". Data are from the 2013 British Attitudes Survey. In the right graph, the variable on the y-axis is the proportion of people in the region that agree with the statement: "When jobs are scarce, men should have more right to a job than women". Data are from the 2005 World Value Survey.

14 years old. The regression controls for ethnicity, survey sample and the level of beliefs elicited in the first question mentioned above. The table shows that men exposed to higher gender occupational segregation tend to have low confidence in their beliefs about men and women's performance in social work.

Table C.1. Correlation between gender occupational segregation and beliefs

DV: Confidence in beliefs of performance in social work		
	(1)	(2)
Online sample:	M	W
Exposure to high gender segregation	-7.149**	2.641
	(3.319)	(3.927)
Observations	110	116
R-squared	0.268	0.169
Mean Dep Var	74.66	80.18

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. The dependent variable is the average of answers to the questions "On a scale from 0 (minimum) to 100 (maximum), how confident are you of your answer [about the performance of a man/woman in social work]?". "Exposure to high gender segregation" is equal to one if the Duncan index of occupational gender segregation in the postcode where a respondent was living when she/he was 14 years old is above the median of the sample. The regression controls for ethnicity, survey wave and the average of the answers to the questions "On a scale from 0 (minimum) to 100 (maximum), what do you think is the performance of a [woman/man] in the social work?". Data are from the auxiliary online surveys and the sample size is determined by the number of people who answered to the postcode question and whose postcode could be matched with the 2011 Census.

D Performance on the job: distributional effects

In this section, I look at the impacts of the treatments on the quality of applicants by measuring changes in the conditional quantiles of applicants' quality. Standard quantile regression models (Koenker and Hallock, 2001) estimate the following conditional quantile function:

$$Q(score_{ia}|X_i) = \alpha + \beta T_i$$

β captures the change in conditional quantile caused by the treatment T_i. For example, suppose that the estimate of β for the 10^{th} percentile of the distribution of standardized test scores is 0.5. This means that an applicant at the 10^{th} percentile of the distribution in the $T_i = 1$ group has a test score that is 0.5 SD higher than an applicant at the 10^{th} percentile of the distribution in the $T_i = 0$ group.

Table D.1. Applicants' skills: quantile regressions

DV: Index of Observable Qualities	(1)	(2)	(3)	(4)	(5)
Quantile	10	30	50	70	90
Women only					
Male Photo	-0.072**	-0.008	0.059**	0.047	0.058
	(0.028)	(0.031)	(0.029)	(0.037)	(0.049)
High Exp Returns	0.002	-0.005	0.008	0.005	-0.000
	(0.029)	(0.031)	(0.029)	(0.036)	(0.046)
Observations	2,062	2,062	2,062	2,062	2,062
R-squared	0.021	0.030	0.032	0.033	0.032
Men only					
Male Photo	0.062	0.097	0.018	0.063	0.197**
	(0.065)	(0.059)	(0.067)	(0.065)	(0.077)
High Exp Returns	0.023	0.120**	0.117*	0.065	0.058
	(0.065)	(0.059)	(0.068)	(0.063)	(0.083)
Observations	440	440	440	440	440
R-squared	0.065	0.067	0.067	0.077	0.078
Basic Controls	✓	✓	✓	✓	✓

Standard errors in parentheses

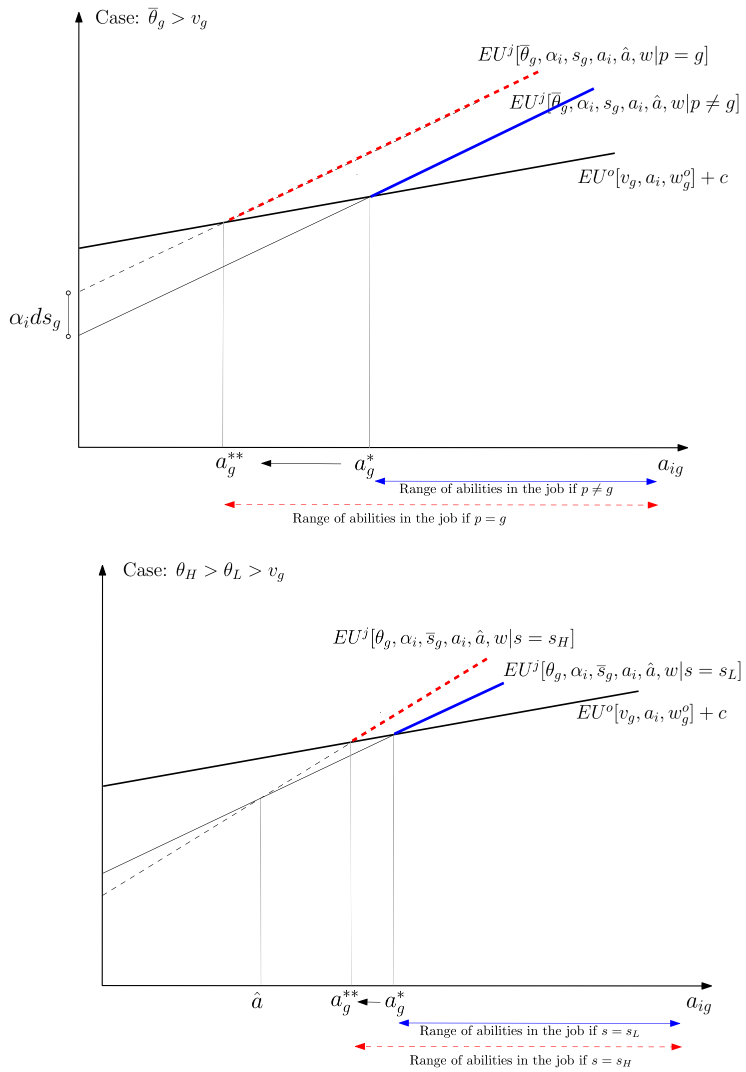
*** p<0.01, ** p<0.05, * p<0.1

Note. Quantile regressions. Estimations are for women in the top panel and for men in the bottom panel. The omitted category is the treatment group that received the female photograph and the low returns information. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment. The regressor "High Exp Returns" is a dummy equal to one for information on high returns to ability treatment. The outcome variable is the "desirable skills index" computed as the mean of the following standardized variables: receiving a first grade, being from a top tier university, frequent past volunteering, high cognitive skills and score in English pre-university tests. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration, non-white ethnicity.

E Appendix to theoretical framework

E.1 Predictions with positive sorting in social work

Figure E.1. Theory predictions: shock to perceived gender shares and expectations of returns to ability in the case of positive sorting in social work

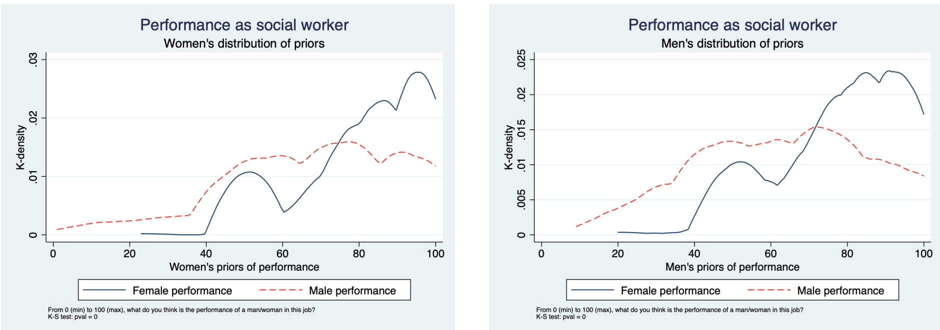


Note. The figure plots the application decision for potential applicants of gender g. Both panels consider the case $U^{j'}(a_i) > U^{o'}(a_i)$. The top panel shows the effect of a shock to perceived gender shares and the bottom panel to expectations of returns to ability. The solid thick line shows expected utility in the outside option.
Top panel: the dashed and thin solid lines show the expected job utility when receiving a photo of the same ($p = g$) or different gender ($p \neq g$), respectively. The vertical distance between these two lines comes from the assumption of the model $E[s_g|p = g] > E[s_g|p \neq g]$. The two thresholds of ability for the marginal applicants a_g^* and a_g^{**} are determined from the intersection of the expected job utility and expected outside option. From Result 1, the size of the applicants' pool is greater when $p = g$ than $p \neq g$ and the marginal applicant a_g^{**} is less skilled than a_g^*.
Bottom panel: the two thin dashed and solid lines show the expected job utility when receiving information on high ($s = s_H$) or low ($s = s_L$) returns to ability. The different slope of these two lines is explained by $E[\theta|s = s_H] > E[\theta|s = s_L]$, as higher returns to ability correspond to a higher slope. The two thresholds of ability for the marginal applicants a_g^* and a_g^{**} are determined from the intersection of the expected job utility and expected outside option. From Result 2, the applicants' pool is larger when $s = s_H$ than $s = s_L$ as long as $B > 0$ and $U^{j'}(a_i) > U^{o'}(a_i)$ and the marginal applicant a_g^{**} is less skilled than a_g^* if $B > 0$.

E.2 Empirical content of the theory assumptions

In this subsection I provide empirical evidence for the assumption of gender differences in priors' average and uncertainty. I use data from the auxiliary online surveys and plot the density of answers to the following question: "On a scale from 0 (minimum) to 100 (maximum), what do you think is the performance of a [WOMAN/MAN] in social work?" where zero is for extremely bad performance, fifty for neither bad nor good performance and a hundred for extremely good performance. The graph on the left-hand side of Figure E.2 shows the distribution of women's beliefs and the one on the right of men's beliefs. Both men and women think that men have on average a lower performance in social work, which leads to the assumption $\theta_M < \theta_W$. The variance of the distribution of beliefs about men is greater than the one of the distribution of beliefs about women, which supports the assumption $\overline{\sigma}^2_M > \overline{\sigma}^2_W$.

Figure E.2. Beliefs about men's and women's performance in social work



Note. Kernel densities of answers to the following question: "On a scale from 0 (minimum) to 100 (maximum), what do you think is the performance of a [WOMAN/MAN] in social work?" The graph on the left-hand side shows the distribution of women's beliefs and the one on the right of men's beliefs. Dashed lines are for beliefs about men's performance and solid lines for beliefs about women's performance.

Table E.1 reports the ten most common past occupations reported in the application form by men and women. As most have had experience in occupations similar to the one they are applying for, the assumption of known a_i seems appropriate.

Table E.1. Most common past occupations for men and women

Men	Women
Social and Community Service Managers	Educational and Vocational Counselors
Child, Family, and School Social Workers	Child, Family, and School Social Workers
Social and Human Service Assistants	Social and Human Service Assistants
Tutors	Tutors
Teacher Assistants	Teacher Assistants
Waiters and Waitresses	Waiters and Waitresses
Personal Care Aides	Childcare Workers
Recreation Workers	Personal Care Aides
Retail Salespersons	Recreation Workers
Customer Service Representatives	Retail Salespersons

Note. Most common past occupations reported in the application form by men and women and converted to standardized SOC4 categories.

E.3 Combining the effects of gender shares and expectations

The assumed additivity between utility from workplace gender composition and expected returns to ability implies that predictions for the four treatment groups follow trivially from results 1 and 2. The following result summarizes these predictions.

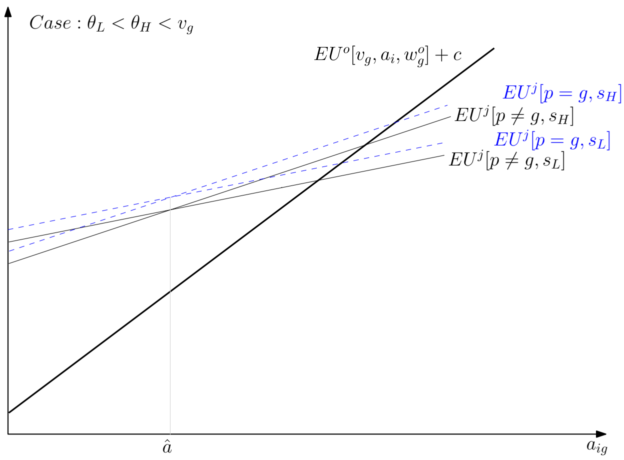
Result 3. ***Interaction between gender shares and expectations***

a. Application rates are highest in treatment $(p = g, s = s_H)$ *and lowest in* $(p \neq g, s = s_L)$

b. Application rates are higher in treatment $(p = g, s = s_L)$ *than* $(p \neq g, s = s_H)$ *iff* $|d\theta_g| < |ds_g|$

Figure E.3 provides the graphical intuition for Result 3 for the case $U^{j'}(a_i) < U^{o'}(a_i)$.

Figure E.3. Theory: gender shares and expectations interacted



Note. The figure plots the application decision for a potential applicant of gender g. The solid black line is the outside option. The two thin solid lines show the expected job utility when receiving information on high $(s = s_H)$ or low $(s = s_L)$ returns to ability and a gender-mismatched photograph $(p \neq g)$. The two dashed blue lines show the expected job utility when receiving information on high $(s = s_H)$ or low $(s = s_L)$ returns to ability and a gender-matched photograph $(p = g)$. The thresholds of ability for the marginal applicants are determined from the intersection of the expected job utility and expected outside option.

E.4 Adding stereotypes to the model

The paper assumes that photographs have no effect on information interpretation. Yet, in the experiment as well as in the real world, the context in which information is conveyed can affect learning and so the photographs manipulation might interfere with people's updating on expected returns to ability on the job.[81]

Work on beliefs in gendered domains (Bordalo et al., 2016, 2019; Coffman et al., 2019) and on confidence (for a review see Bertrand, 2011) points to an interaction between job difficulty and gender-specific expectations. Bordalo et al. (2019) find that bringing gender comparisons top of mind affects people's beliefs of own ability across domains: women paired with men, relative to women paired

[81] There is rich experimental evidence on people's "mental gaps" in information gathering and processing (Handel and Schwartzstein, 2018).

with women, become more optimistic about own performance as female advantage increases. The evidence I show on women's behavior is consistent with a model in which women's estimation of own performance is decreasing when paired with men in a challenging task, but increasing when paired with women in the same task. Thus it's possible that the male photograph makes women revise their gender advantage in the job. Along these lines, I propose a simple learning mechanism through which gender shares might affect updating of returns to ability on the job. Suppose that individual ability a_i is the sum of a mean-zero individual component, ϕ_i, and a gender comparison component $a_g^{st} = a_g - a_{-g} : a_i = a_g^{st} + \phi_i$. In a female-dominated job, stereotypes imply $a_W^{st} > 0 > a_M^{st}$.

By changing the gender composition in the job, photographs might affect beliefs on a_g^{st}. For women, own gender advantage is smaller when there is a higher male proportion in the job (as inferred by seeing a male photograph).[82] What's bad news for women is good news for men: seeing a male photograph could positively affect a_M^{st} and reduce their gender disadvantage. This modelling assumption is equivalent to assuming that parameter $\hat{a}$ is a function of s_g. Assumption 3 formalizes this.

Assumption 3. ***Gender stereotypes***
$\forall g \in \{W, M\} : E[\hat{a}|p = g] < E[\hat{a}|p \neq g]$

Adding stereotypes to the model makes ambiguous the predictions on the interaction between treatments. In particular, in this model application rates are not necessarily the lowest in treatment $(p \neq g, s = s_L)$, but can be lowest in treatment $(p \neq g, s = s_H)$. Let's take a simple case for the sake of explanation and focus on women's results. Following assumption 3, $\hat{a}$ is greater in the male photograph than in the female photograph treatment. If this difference is big enough, it can lead to a situation in which the marginal applicant has ability higher than $\hat{a}$ in the female photograph treatment and lower than $\hat{a}$ in the male photograph treatment (i.e. condition B, $a_i^* > \hat{a}$, is satisfied only in one treatment). This implies, in turn, that the difference in application rates between receiving information on high versus low returns to ability is positive conditional on a female photograph and negative conditional on a male photograph. This is exactly what I find on women in the data.

E.5 Proofs

Proof. **Existence of threshold of ability a_i^***
Define $U^j(a_i) = U^j(a_i, \hat{a}, s_g, \alpha_i, \theta_g)$ and $U^o(a_i) = U^o(a_i, c, v_g, \bar{w})$. Consider a closed intervals of ability a_i: $[a_1, a_2]$, with a_1 and a_2 bounded away from 0 and infinite. Assume that $U^j(a_i)$ and $U^o(a_i)$ satisfy the following conditions:

a0. They are both continuous in the interval $[a_1, a_2]$

a1. $U^j(a_1) < U^o(a_1)$

a2. $U^j(a_2) > U^o(a_2)$

[82] Notice that, in a partial equilibrium framework in which men and women's abilities are given, this is inconsistent with the evidence shown in Figure E.2. Both men and women think that men are worse in social work than women. Thus a higher proportion of men in the job should imply a lower aggregate performance and a bigger advantage for women that enter. However, in a general equilibrium framework, a higher proportion of men in the job might signal that they are actually better than previously thought, leading to the hypothesized effect.

Define the function $H(a_i) = U^j(a_i) - U^o(a_i)$, which is continuous as well in $[a_1, a_2]$. Then:

$H(a_1) = U^j(a_1) - U^o(a_1) < 0$ from a1

$H(a_1) = U^j(a_2) - U^o(a_2) > 0$ from a2

Since $H(.)$ is continuous, by the Intermediate Value Theorem (IVT) there must be a value $a_i^* \in [a_1, a_2]$ such that $H(a_i^*) = 0$. Thus the two functions $U^j(a_i)$ and $U^o(a_i)$ must intersect in a_i^*. In the application decision for the marginal applicant, if the minimum value of a_1 is zero, the IVT conditions imply that $\theta_g > \frac{\alpha_i s_g - \bar{w} - c}{\hat{a}}$ and $\theta_g > v_g$. □

Proof. **Result 1**

We need to consider how the change in own gender proportion s_g affects the marginal applicant's ability. Define $G(a_i, \hat{a}, s_g, \alpha_i, \theta_g, c, v_g, \bar{w}) = U^j(a_i) - U^o(a_i)$, where $U^j(a_i)$ and $U^o(a_i)$ are as defined in the previous proof. Consider the vector $\overline{x}_0 = (a_{i0}, \hat{a}_0, \alpha_{i0}, s_{g0}, \theta_{g0}, \bar{w}_0, c_0, v_{g0})$ such that $G(\overline{x}_0) = 0$. Assume that $\frac{\partial G(\overline{x}_0)}{\partial a_i} \neq 0$. By the Implicit Function Theorem (IFT):

$$\frac{\partial a_i}{\partial s_g} = -\frac{\frac{\partial G(.)}{\partial s_g}}{\frac{\partial G(.)}{\partial a_i}}$$

From the definition of $G(.)$:

- $\frac{\partial G(.)}{\partial s_g} = \frac{\partial U^j(.)}{\partial s_g} = \alpha_i$. Thus $sign\left(\frac{\partial G(.)}{\partial s_g}\right) = sign(\alpha_i) > 0$ under the assumptions of the model.
- $\frac{\partial G(.)}{\partial a_i} = \frac{\partial U^j(.)}{\partial a_i} - \frac{\partial U^o(.)}{\partial a_i} = \theta_g - v_g$. The sign of this difference depends on the relative slope of the on-the-job expected utility and the outside option.

It follows that $sign\left(\frac{\partial a_i}{\partial s_g}\right) = -sign\left(\theta_g - v_g\right)$. This implies that a decrease in perceived own gender proportions s_g will decrease (increase) the marginal applicant's ability a^* if the best (worst) people select into the job. In both cases, there is an increase in the mass of people applying to the job. The magnitude of the change in a^* is independent of a^* level, increasing in α_i and decreasing in $v_g - \theta_g$. □

Proof. **Result 2**

We need to consider how the change in expected returns to ability θ_g affects the marginal applicant's ability. Consider $G(a_i, \hat{a}, s_g, \alpha_i, \theta_g, c, v_g, \bar{w}) = U^j(a_i) - U^o(a_i)$ as defined in the previous proof. Consider the vector $\overline{x}_0 = (a_{i0}, \hat{a}_0, \alpha_{i0}, s_{g0}, \theta_{g0}, \bar{w}_0, c_0, v_{g0})$ such that $G(\overline{x}_0) = 0$. Assume that $\frac{\partial G(\overline{x}_0)}{\partial a_i} \neq 0$. By the Implicit Function Theorem (IFT): $\frac{\partial a_i}{\partial \theta_g} = -\frac{\frac{\partial G(.)}{\partial \theta_g}}{\frac{\partial G(.)}{\partial a_i}}$.

From the definition of $G(.)$:

- $\frac{\partial G(.)}{\partial \theta_g} = \frac{\partial U^j(.)}{\partial \theta_g} = a_i - \hat{a}$. Thus $sign\left(\frac{\partial G(.)}{\partial s_g}\right)\Big|_{a_i^*} = sign(a_i^* - \hat{a})$. Solving for a_i^*, this implies the condition on the sign of B: $a_i^* > \hat{a}$ if $\bar{w} + c - \alpha_i s_g + v_g \hat{a} < 0$ (or equivalently $B > 0$).
- $\frac{\partial G(.)}{\partial a_i} = \frac{\partial U^j(.)}{\partial a_i} - \frac{\partial U^o(.)}{\partial a_i} = \theta_g - v_g$. The sign of this difference depends on the relative slope of the on-the-job expected utility and the outside option.

It follows that there are four possible cases for $sign\left(\frac{\partial a_i}{\partial \theta_g}\right)$, given by the combination of one level of a_i^* - above or below $\hat{a}$ - and the relationship between on-the-job and outside option returns to ability. These cases are summarised in the Table below. A positive sign of the derivative of a_i with respect to θ_g means that we expect an increase in the number of applications when on-the-job marginal returns increase. From the cross derivative of a_i wrt θ_g and a_i, the magnitude of the change in a^* is proportional to $|\theta_g - v_g|$.

	$\theta_g - v_g > 0$		$\theta_g - v_g < 0$	
	$a_i^* > \hat{a}$	$a_i^* < \hat{a}$	$a_i^* > \hat{a}$	$a_i^* < \hat{a}$
$\frac{\partial a_i}{\partial \theta_g}$	$-$	$+$	$+$	$-$

□

Proof. **Result 3: interaction between gender shares and expectations**

To understand the total effect of receiving a signal s and a contemporaneous change in perceived gender proportions, I compute the total differential $a_i|_{a_i^*}$:

$$da_i|_{a_i^*} = \frac{\partial a_i}{\partial \theta_g} d\theta_g + \frac{\partial a_i}{\partial s_g} ds_g$$

The proof entails the comparison of the total differential between each pair of the four treatment groups. Comparing two emails with the same photograph (statistic) implies $ds_g = 0$ ($d\theta_g = 0$), thus results 1. and 2. apply. The crucial comparison is between treatments with both different photographs and statistics: (g, θ_H) vs $(-g, \theta_L)$ and $(-g, \theta_H)$ vs (g, θ_L). Let's consider the first case (the same reasoning applies to the second).

Comparing (g, θ_H) vs $(-g, \theta_L)$ means that $ds_g > 0$ and $d\theta_g > 0$. If $B > 0$ and $\forall sign(\theta_g - v_g)$, $sign(\frac{\partial a_i}{\partial \theta_g}) = sign(\frac{\partial a_i}{\partial s_g})$, thus the two changes reinforce each other. This will implies that in absolute value the total change in a_i, at the margin, is biggest between treatments (g, θ_H) and $(-g, \theta_L)$. Thus the marginal applicant's ability will be maximum in treatment (g, θ_H) and minimum in treatment $(-g, \theta_L)$ when $\theta_g - v_g < 0$. If $B < 0$, the sign of this comparison is instead ambiguous. If $\theta_g - v_g < 0$:

$$da_i|_{a_i^*} = \underbrace{\frac{\partial a_i}{\partial \theta_g}}_{-} \underbrace{d\theta_g}_{+} + \underbrace{\frac{\partial a_i}{\partial s_g}}_{+} \underbrace{ds_g}_{+}$$

The sign of the total differential depends on the relative strength of the change in marginal returns to ability and the change in gender proportions. If $|d\theta_g| > |ds_g|$, then the change in expected returns to ability prevails and marginal ability decreases, counteracting the positive change generated by the photograph.

□

F An additional exercise on overconfidence

In this section I further explore whether the effect of the information treatment can be explained by (gender differences in) overconfidence. I use the survey questions defined at the end of Section C to construct a proxy of individual over-precision in their priors on men and women's performance in female jobs. I select the most important observable predictors of this measure using Lasso regression and impute the coefficients to my experimental sample. This provides a measure of "predicted confidence" (overprecision) in others' performance in social work and teaching.

Table F.1 shows the treatment effect on men's application likelihood depending on their predicted confidence. The increase in application rates is driven by men with over-precision below the median. As long as this is correlated with a higher likelihood of under-placement of own ability with respect to others, it suggests that the effects are actually driven by the least confident men.[83] Moore and Healy (2008) show that lack of precision on beliefs about others is positively correlated with overplacement in easy tasks, but also positively correlated with under-placement in hard tasks. In other words, unprecise estimates of others' performance increase people's tendency to under-place one's own performance in hard tasks. This seems the relevant case in my context.

Table F.1. Treatment effects by predicted priors' uncertainty

	DV: Applied and never D.O.			
	(1)	(2)	(3)	(4)
	Confidence in women's ability		Confidence in men's ability	
	< med	> med	< med	> med
High Exp Returns	0.108**	0.041	0.137***	0.017
	(0.048)	(0.049)	(0.049)	(0.051)
Observations	394	398	386	406
R-squared	0.024	0.016	0.025	0.014
Mean Dep Var	0.53	0.56	0.48	0.62

Bootstrapped se in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates for men only. Columns (1) and (2) split the sample at the median level of predicted confidence in priors about women's performance in social work and primary school teaching. Columns (3) and (4) do the same for priors about men. The variables used to predict confidence are age, whether the person studied in a top university, non-white ethnicity, whether the person studied a subject aligned with the job, exposure to occupational gender segregation and gender. The omitted category is the group that received information on low expected returns to ability.

[83] This table is also consistent with the hypothesis that information provision benefits the most men who start off with greater uncertainty about returns in female-dominated jobs.

G The trade-off between retention and performance: a survey

In October 2020, I sent a survey to the management team of the partner organization to ask about their opinion regarding the trade-off between performance and retention in the programme (N=31). Half of the respondents are supervisors of social workers in local communities and half belong to recruitment functions. 55% of the respondents have been working in the organization for less than two years and 45% of them have managerial responsibilities. I used two vignette-type questions (see Figure G.1). For both vignettes, one scenario is such that all the people hired stay for the full two-year programme, but average performance in practice tests is moderate. In the alternative scenario, 18% of people exit before finishing the programme, but there is a higher average performance net of leavers. I ask respondents to choose their favourite scenario i) with just this basic information (question 1) and ii) adding information about gender diversity in the two scenarios (question 2). I also elicited the performance level that would make them indifferent between the two scenarios after their choice. The statistics used in the two scenarios come from data on women from treatment groups (W,L) and (M, H).

> Consider the following two scenarios. Consider them identical besides the characteristics given below. Which of these scenarios would you prefer for the local communities where the organization operates? In both, consider as top performer somebody with commendable or excellent average scores ($\geq 60\%$) in the practice tests of the programme.

Figure G.1. Survey questions

(a) Vignette question 1

Scenario A	Scenario B
■ Starting cohort of 100 people	■ Starting cohort of 100 people
■ **Everyone** completes the programme (2 full years)	■ **18 people** drop-out before finishing the programme
■ **51%** are top performers by the end of the programme	■ **61%** are top performers by the end of the programme

(b) Vignette question 2

Scenario A	Scenario B
■ Starting cohort of 100 people	■ Starting cohort of 100 people
■ **Everyone** completes the programme (2 full years)	■ **18 people** drop-out before finishing the programme
■ **51%** are top performers by the end of the programme	■ **61%** are top performers by the end of the programme
■ Women are **85%** by the end of the programme	■ Women are **71%** by the end of the programme

Note. The figure shows the two main questions asked in the survey with the organization about the trade-off between retention and performance.

[Online Appendix]

Online Appendix

A Outside option: methodology

I compute the individual current expected hourly wage in the U.K. as a measure of the individual outside option. Using the U.K. Labor Force Survey (LFS) quarterly data between January 2017 and December 2018, I estimate a Mincerian regression of the log-hourly wage on a set of observables which are available both in the LFS and my experimental dataset.[84] I then impute the coefficients of the Mincerian regression to my experimental data to predict an individual-level expected wage in the UK labour market. I interpret this measure as the individual outside option component w^o.[85]

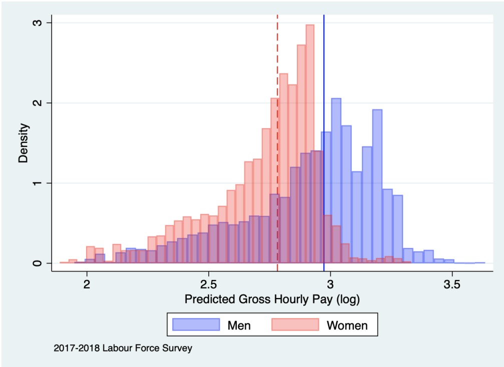
I limit the sample to men and women between 16 and 64 years old and, to match the eligibility criteria of my organization, I include only people who have at least a bachelor degree or, if students, who are currently studying towards a bachelor degree or higher university title. Following the LFS guidance, the variable for the hourly pay has been truncated between 0 and 99 (variable called HOURPAY) and has been derived from the variables GRSSWK (gross weekly pay), POTHR (usual hours of paid overtime) and BUSHR (usual hours worked in main job, excluding overtime). I first take the natural logarithm of the HOURPAY variable before running the regression. The hourly pay is computed for all respondents who are employees and those on a government scheme. I estimate a Mincerian regression of the log-hourly wage on the following set of dummies: university subject (16 categories), age, age squared, British nationality, gender, marital status, non-white ethnicity, first grade in university. I do not control for the fulltime employment status because the coefficient would bias upward the estimated outside option of people in fulltime employment as compared to students in my sample. Figure OA.1 shows the distribution of the computed outside option by gender. Table OA.2 shows the coefficients of the Mincerian regression on the LFS data. The omitted category are non-married white women who studied Arts.

Table OA.1 compares a random subsample from the LFS with the experimental sample. I generated the former to reproduce the same age distribution of the latter. Both men and women in my experiment are more likely to be of non-white ethnicity, less likely to be married, less likely to have graduated before 2016, more likely to have worked in the public sector or healthcare and, relatedly, less likely to have studied scientific subjects. These differences confirm that people in the experimental sample are selected on the basis of greater interest in public sector and/or healthcare jobs.

[84] For more information on the Labour Force Survey, see the LFS website: discover.ukdataservice.ac.uk/series/?sn=2000026.

[85] A drawback of this measure is that it rewards experience and other observable demographics over talent, whose only measure in both the LFS and my data is university grade. This means that it might overestimate the opportunities available to older and less skilled people as compared to younger more skilled ones.

Figure OA.1. Outside option distribution by gender



Note. The figure shows the distribution of outside option for men (in blue) and women (in red). The red dashed (blue solid) line is the women's (men's) median.

Table OA.1. Labour Force Survey and experimental sample comparison

	Labour Force Survey					Experiment	
	Women		Men		Diff (1)-(2)	W	M
	Mean	SD	Mean	SD	p-val	Mean	Mean
Non-white	.12	.33	.14	.34	.07	0.27	0.28
Age	28.77	8.36	29.3	8.73	.01	26.35	28.68
Married	.28	.45	.27	.44	.51	0.12	0.19
First Grade	.15	.35	.14	.35	.3	0.18	0.20
Graduated before 2016	.73	.44	.75	.44	.19	0.34	0.45
FTE in Public Sector	.49	.5	.27	.44	0	0.71	0.60
Scientific Subject	.15	.36	.32	.47	0	0.05	0.09
Aligned Subject	.44	.5	.27	.45	0	0.70	0.48

Note. The first five Columns of the table show summary statistics from a random sample of the LFS which I generated to reproduce the same age distribution of the experimental sample. The last two Columns show the shares of women and men with the given row characteristic in the field experiment. Column "Diff (1)-(2)" contains the difference in the proportions of women and men that have the characteristic of the corresponding row in the LFS sample. "FTE in Public Sector" is an indicator variable for working in the government and includes jobs in healthcare.

Table OA.2. Mincerian regression to predict outside option

DV: Log Hourly Pay			
Other or missing	0.0634***	Architecture	0.204***
	(0.017)		(0.028)
Medicine	0.518***	Social Studies	0.195***
	(0.029)		(0.019)
Allied to medicine	0.121***	Law	0.267***
	(0.018)		(0.023)
Biology	0.141***	Business	0.216***
	(0.019)		(0.018)
Agriculture	0.106***	Communications	0.0677***
	(0.030)		(0.025)
Physics	0.211***	Languages	0.122***
	(0.021)		(0.023)
Maths and IT	0.282***	History	0.101***
	(0.020)		(0.024)
Engineering	0.318***	Education	0.136***
	(0.019)		(0.018)
Age	0.0935***	Male	0.143***
	(0.002)		(0.007)
Age squared	-0.001***	British	0.0173
	(0.000)		(0.012)
Married	0.0889***	Non-white	-0.0533***
	(0.008)		(0.012)
First Grade	0.0954***	Constant	0.467***
	(0.011)		(0.048)
Observations		22325	
R-squared		0.235	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

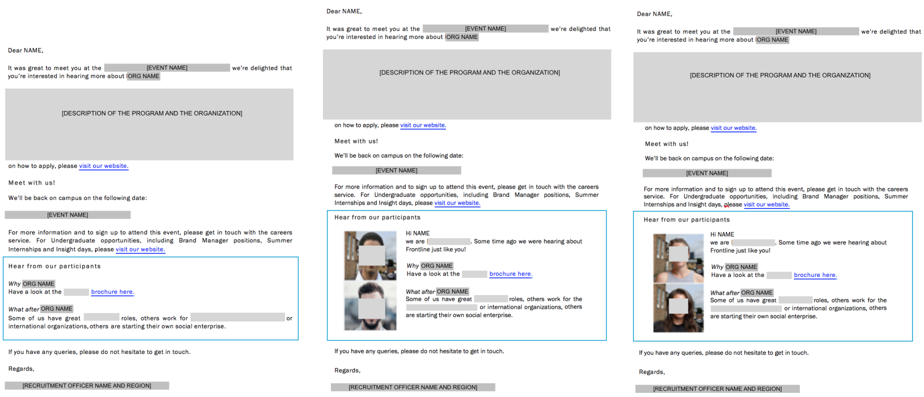
Note. OLS regression. The table reports the coefficients from a regression of log hourly wage on seventeen university subject categories, age, age squared, gender, marital status, ethnicity, British citizenship and having achieved a first grade in university. The omitted category are non-married white women who studied arts.

B External validity of the gender shares effect: evidence from a complementary experiment.

I address external validity of the null result of the photograph manipulation through a complementary field experiment with the same partner organization. The goal is to understand the extent to which gender shares affect men's decision to apply for a female-dominated job in a sample which is less selected on interest in the job.

Between September and November 2017 the partner organization visited 52 universities across the country conducting a variety of career events (e.g., stands at job fairs, workshops, presentations). The main goals of these events are to promote the organization's program and encourage applications. On average, each university was visited slightly more than three times, for a maximum of six. Each university is assigned to a Recruitment Officer (RO) who is in charge of organizing and conducting the events, collecting email addresses of event participants and sending a follow-up email with further information about the program.[86]

Figure OB.1. Experiment in universities: treatments



People who took part to career events and left their email address in a mailing list were randomly assigned to three groups, which differed in the format of the follow-up email received. The text content of these three emails was exactly the same, but they might show i) no picture, ii) a picture of previous female workers, ii) or a picture of previous male workers. The three email templates are shown in Figure OB.1. Assignment to treatment was stratified by university, event and gender.

Each email contains links to the organization's website which are trackable at the level of stratification and treatment. This allows me to know the number of participants of gender g in event e in university u that clicked on any email link, whether they are first time users and some metrics of their online behaviour for each treatment group.[87] The main outcome of this experiment is whether people click on "Apply" on the organization's website. Each event had an average number of 30 sign-ups,

[86] RO's performance evaluation does not depend on the number of email addresses collected at university events. Mailing lists were collected in 75% of the total number of events run by the organization. Out of the remaining 25%, ROs couldn't collect participants' email addresses for three main reasons: i) time constraints, ii) the university refused to share participants' data or iii) all the participants had already signed-up. In two events the email lists were collected, but the RO just sent a standard follow-up email template.

[87] Online behaviour is measured using standard metrics recorded by Google Analytics.

for a total of 2877 unique participants. Table OB.1 presents summary statistics of the experimental sample and balance checks. 78% of participants are last year students or graduates and 21% of them are or were enrolled in a science or business course. Overall, 29% of the event participants have heard about the organization before, mostly through news and ads. Men represent 22% of the sample, for a total of 630. At baseline, men are less likely to access the organization website as compared to women: on average, only 2% of men click on any link as compared to 9% of women.

Table OB.1. Experiment in universities: balance and summary statistics

	Overall			Joint test		Pairwise tests	
	N	*mean*	*sd*	*F stat*	*p-value*	*min p value*	*max diff*
Male	2877	0.22	0.42	1.397	0.248	*0.095	0.032
Last year	2500	0.58	0.49	0.120	0.887	0.662	-0.011
Graduates	2500	0.10	0.30	0.298	0.742	0.453	0.011
First/second year	2500	0.32	0.47	0.067	0.935	0.739	-0.008
Science or business	2334	0.21	0.41	1.230	0.292	0.168	0.028
Heard about the job	2334	0.29	0.45	0.863	0.422	0.245	0.027
- on campus	1221	0.21	0.41	1.411	0.244	0.125	0.043
- in news/ads	1221	0.55	0.50	1.492	0.225	*0.091	-0.058
- from friends	1221	0.07	0.26	0.090	0.914	0.680	0.008
- online	1221	0.17	0.37	0.317	0.729	0.454	-0.020

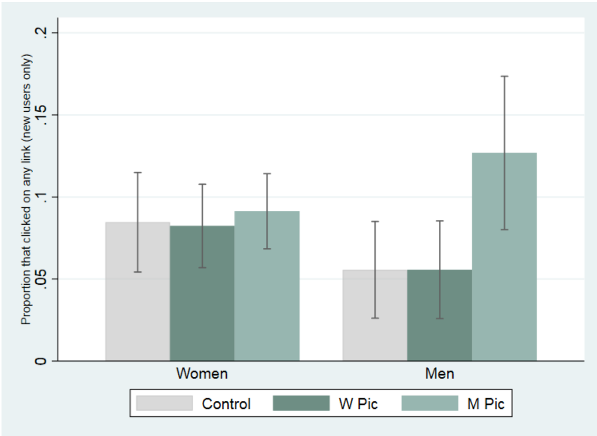
Note. "Last year" and "First/second year" are indicator variables for the year of enrolment in university. "Science or business" is an indicator for studying a scientific or economics/business subject. "Heard about the job" is equal to one if the person heard of the organization before attending the event. Columns 4 and 5 report the F-statistic and p-value from a joint test of the significance of the set of treatment dummies in explaining each row variable with robust standard errors. The last two Columns report the minimum p-value and maximum difference from t-tests between pairs of treatment groups.

Results indicate that men are more likely to access the organization's website when they receive an email containing male photographs as compared to both the control group and the treatment with female photographs. Figure OB.2 shows that the number of clicks by men almost doubles in the male photograph treatment. However, men's behaviour on the website does not translate into more applications. Table OB.2 estimates the effect of each of the treatment emails on application for people of gender group g, event e and university u using the following specification:

$$y_{geu} = c + \beta_1 MPic_{geu} + \beta_2 WPic_{geu} + X'_{eu}\beta_3 + \delta_u + \epsilon_{geu}$$

The regressions include university fixed effects δ_u and the vector of event controls X_{eu} (type of event, month, number of participants, gender of RO). I use robust standard errors as the randomization was at the individual level and add analytical weights by treatment group size. Table OB.2 shows that the male photograph treatment doesn't increase men's applications, which reinforces the external validity of the null effect of the male photograph.

Figure OB.2. Experiment in universities: results



Note: The bar chart shows the proportion of clicks by new users in the different treatment groups of the experiment.

Table OB.2. Experiment in universities: effects on applications

DV: Event participant registered to apply		
	(1)	(2)
VARIABLES	M	W
Women's Pic	-0.046	0.017
	(0.038)	(0.028)
Men's Pic	0.008	0.007
	(0.054)	(0.029)
Scientific Subject	-0.075**	-0.107***
	(0.029)	(0.024)
Observations	337	1,259
R-squared	0.148	0.109
Mean Dep Var	0.082	0.17

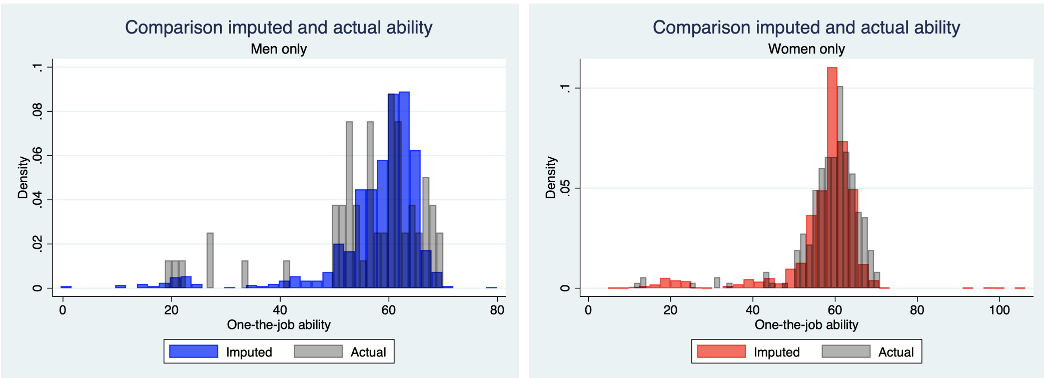
Clustered standard errors in parentheses (uni level)
*** p<0.01, ** p<0.05, * p<0.1

Note. OLS regressions for men and women separately. The dependent variable is equal to one of the participant filled-in the online registration form necessary to apply for the job. The omitted category is the group receiving emails with no workers' photographs. "Women's Pic" and "Men's Pic" are indicator variables for the two experimental treatments. The regression includes university fixed effects and event controls X_{eu} for event type, month, number of participants and gender of RO. I add analytical weights by treatment group size. The table limits the sample to last year students or graduates.

C Logit estimation: methodology

Section “Heterogeneity by job-specific ability” in the paper uses a proxy for job-specific ability which is the predicted on-the-job performance score, obtained from the pure control group through a linear truncated regression using the following regressors: ranking and average completion rate of the university attended by the candidate, subject studied, obtaining a first grade, whether the grade is expected or obtained, age, age squared and whether the person is in FTE. The following figure shows the distribution of imputed ability against the distribution of actual test scores in the job from the raw data, by gender. The following table shows the coefficients of the logit estimation.

Figure OC.1. Comparison of imputed and actual on-the-job performance



Note. The figure shows the comparison of imputed and actual on-the job performance distributions. The histograms on the left-hand side are for men and on the right-hand side for women. Ability is on a scale from 0 (min) to 100 (max).

Table OC.1. Logit estimation: output by gender

DV:	Applied and never DO	
	(1)	(2)
	Men	Women
$w - \overline{w}$	0.171	0.736**
	(0.253)	(0.319)
a_i	0.006	-0.011
	(0.013)	(0.011)
Treat θ_H	0.303**	-0.070
	(0.144)	(0.148)
Treat $\theta_H \times a_i$	0.018	0.009
	(0.017)	(0.015)
Treat $p = g$	-0.059	0.215
	(0.146)	(0.147)
Constant	-0.010	0.229*
	(0.133)	(0.134)
Observations	807	3,513

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

D Heterogeneous effects by outside option parameters

Table OD.1. Treatment effects by regional wage dispersion

DV: Applied and never DO				
	(1)	(2)	(3)	(4)
	Men		Women	
	Wage dispersion:		Wage dispersion:	
	Low	High	Low	High
Male Photo	0.004	-0.075	-0.055***	-0.045
	(0.042)	(0.061)	(0.020)	(0.030)
High Exp Returns	0.058	0.113*	-0.007	-0.032
	(0.042)	(0.062)	(0.020)	(0.030)
Observations	555	252	2,449	1,064
R-squared	0.014	0.065	0.018	0.007
Basic controls	Y	Y	Y	Y
Mean Dep Var	0.51	0.54	0.60	0.58
Photo = Exp Ret p-val	0.37	0.03	0.09	0.77
Rand Inf p-val				
Photo	0.91	0.24	0.007	0.13
Exp Returns	0.18	0.07	0.75	0.30

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note. OLS estimates run separately for men (Columns 1 and 2) and women (Columns 3 and 4). “Wage dispersion” is computed as the 75/25 interquartile range of the gender-specific distribution of hourly wages across industries in the UK region where the candidate lives. The level “high” or “low” is defined for values of the index respectively above or below the gender-specific median in the experimental sample. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration and non-white ethnicity. The rows “Rand Inf p-val” contain the p-values of the coefficients on the treatment dummies from randomization inference (randomization-t) with 1000 repetitions.

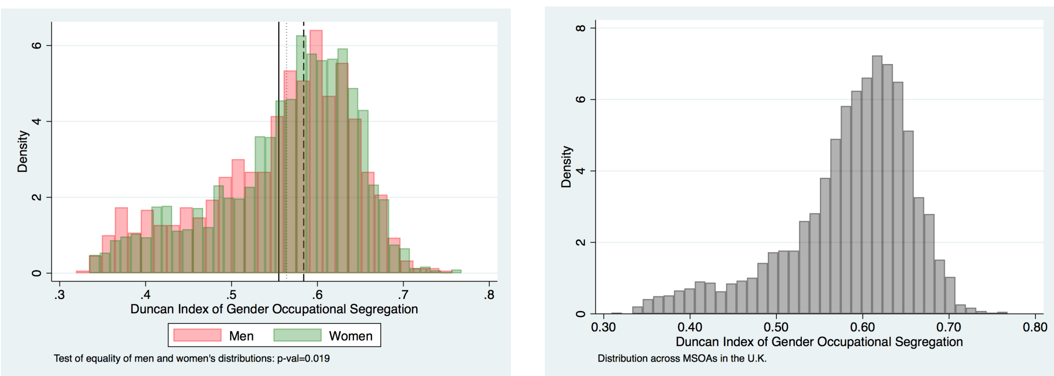
E Heterogeneous effects by occupational gender segregation

I use microdata on the local occupational structure by gender from the 2011 U.K. Census to construct the Duncan index of occupational segregation (Duncan, 1955). The dataset contains the distribution of workers by gender across 362 detailed SOC4 occupational categories at the Medium Layer Super Output Areas (MSOA) level. The sample is a 10% random sample from the 2011 Census. In 2011, there were 7201 MSOA in the UK and the median MSOA comprised 188 8-digits postcodes, with a minimum of 89 postcodes to a maximum of 1033.

The Duncan index is computed using the following formula: $\frac{1}{2}\sum_{i=1}^{N}|\frac{m_i}{M} - \frac{f_i}{F}|$, where m_i and f_i are the male and female population, respectively, in occupation i and M and F are the total working population in the local labour market. The index takes values between 0 (complete integration) and 1 (complete segregation) and identifies the percentage of women (or men) that would have to change occupations for the distribution of the two genders to be equal.

Using a bridge between the Census local area codes and 7-digit postcodes, I merged the indexes with my experimental data through the subjects' secondary school postcode and, when missing (for 62% of subjects), home postcode. The use of the secondary school postcode is motivated in the main body of the paper. The subsample of subjects with only home postcode available is made of 50% students and 50% workers. For students, home postcode is mostly the postcode of their parents' home, which is likely to be where they grew up. For workers, it is instead the current domicile. The distribution of the Duncan index in my experimental sample is representative of the overall Country, as shown in Figure OE.1. The U.K. average Duncan Index across MSOAs is 0.5839 and the average in my sample is 0.563.

Figure OE.1. Duncan Index in the experimental sample and in the UK



The figure on the left shows the distribution of the Duncan Index in the experimental sample by gender (postcode level). We can see that men's distribution is shifted to the left of women's distribution (Kolmogorov-Smirnov test of equality of distributions: p-val=0.019). The vertical black line shows the mean for men (0.554) and the vertical dashed line shows the mean for women (0.564). The distribution for the whole U.K is showed in the figure on the right (MSOA level).

I use the Duncan Index as an individual measure of exposure to gender-segregated labour markets in the previous decade before the current job application. One shortcoming of this method is that it does not equalize the age of exposure to local labour markets across candidates. Timing of exposure has been shown to be a crucial variable for norms internalization (Heckman and Kautz, 2012). This implies that the Duncan index computed using data from 2011 is likely to be weakly correlated with gender norms for people who were older than 23 at the moment of application. But the Duncan index showed little change over the last two decades (Blau et al., 2013) and the correlation in my experimental data between the 2001 and 2011 Duncan index is 0.70 (p-val = 0.000). Nevertheless, the results of Table 3 of the paper are robust to assigning the Duncan index computed from the 2001 Census data to individuals older than 23 (60% of men's sample).