Breaking Gender Barriers: Bringing Men into the Pink-Collar Jobs of the Future

Alexia Delfino*

November 27, 2019

JOB MARKET PAPER

Click here for the most recent version

Abstract

Traditionally female-dominated sectors are growing while traditionally male-dominated sectors are shrinking. And yet, sectorial male shares are not changing accordingly. Why don't men enter female-dominated occupations? I study men's selection into social work, a fast-growing occupation where the share of men has historically been below 25 percent. I embed a field experiment in the UK-wide recruitment of social workers to analyse barriers to men's entry and the nature of men's sorting into this occupation. 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 application decisions, which suggests no role for gender identity or social stigma in their choices. Increasing expected returns to ability encourages men to apply, and improves the average quality of the applicants and performance on the job of the new hires, indicating that men are negatively sorted into social work. In turn, a higher (perceived) share of male workers improves the quality of female hires by discouraging the least talented women from applying. These findings suggest that breaking barriers to men's entry in female-dominated occupations may help employers increase the diversity and overall quality of their workforce.

---

*London School of Economics, a.delfino2@lse.ac.uk. I thank the Director of Selection and all the members of the Attraction and Selection teams in the partner organization. This project would have not been possible without them. I thank Nava Ashraf, Oriana Bandiera, Matthew Levy and Erik Eyster for their priceless advice throughout the project. I also thank useful advice from Robin Burgess, Maitreesh Ghatak, Gharad Bryan, Greg Fischer, Rachel Kranton, Marco Manacorda, Barbara Petrongolo, Christine Exley, Fabio Paglieri, Erik Hurst, Catherine Thomas, Jordi Blanes i Vidal, Eliana La Ferrara, David Deming, Abhijit Banerjee, Robert Akerlof, Alwyn Young, Guy Michaels, Steve Pischke, Andrew Ellis, Sandra McNally, Gilat Levy, Yona Rubinstein, Florian Ederer, Florian Englmaier, Emir Kamenica, Matthew Lowe, Johanna Rickne, Alessandra Voena, Abi Adams, Natalie Bau, Corinne Low, Erina Ytsma. Celine Zipfel, Kim Sarnoff, Francesco Sannino, Matteo Benetton, Michel Azulai, Nicola Limodio, Andreas Ek, Giulia Giupponi, Vincenzo Scrutinio, Clement Minaudier and Tiancheng Sun 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

The shift from brawn-intensive to brain-intensive occupations has decreased 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). Female-dominated industries, such as health and education, displayed the highest growth, and yet their gender composition barely changed despite falling male economic activity (Blau and Kahn, 2017). Understanding the barriers to men's entry in these occupations is important to move workers in declining industries towards new opportunities.

In this paper, I study men's selection into one of such high-growth female-dominated occupations: social work. Over the next decade, the growth rate of social workers is expected to be twice the average growth of US occupations (Bureau of Labor Statistics, 2019), but the male share of social workers has not changed since 1970 (Blau et al., 1998). This can be the result of men not applying or employers not selecting those who apply. Understanding the nature of sorting is crucial to design tools that increase diversity without lowering the quality of the expanding workforce in this sector.

Guided by a theoretical framework, I explore the barriers to men's entry and the nature of their sorting into female-dominated jobs by generating experimental variation in the recruitment strategy for a real job in social work and then following what happens to applicants 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 applications.

I embed a field experiment in the UK nationwide recruitment of social workers to exogenously vary two key determinants of selection: perceived gender shares and expectations of returns to ability. The former embodies non-pecuniary factors related to the association between an occupation and a certain gender, which have been shown to be relevant in labour supply decisions (Akerlof and Kranton, 2000), and the latter represents standard incentives in occupational choice. Disentangling these two channels in observational data is difficult as it requires two independent sources of exogenous variation.[1] A controlled setting allows me to overcome this identification challenge, but might still affect participants choices through novelty or experimenter effects.

I overcome these risks by working in collaboration with one of the main organizations in the sector and introducing variation in the content of recruitment messages sent to their potential applicants.[2] By conducting the experiment in 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. I can thus check whether more applications lead to more and better hires. Compared to alternative sources of variation, for instance monetary incentives, my design preserves the organizational systems in place, is easily scalable and mimics common low-cost policies that employers use to increase gender diversity.[3]

[1] This is analogous to the empirical challenge of distinguishing preference-based from inference-based discrimination (Altonji and Blank, 1999; Guryan and Charles, 2013; Neumark, 2018).

[2] Potential applicants need to register their interest in applying on the website of the partner organization. This implies that the experimental sample is selected on the basis of a minimum level of interest in the job. First, a minimum interest in the job makes this the relevant sample from a policy perspective. Moreover, the brevity of the form and the application rate after registration (between 50% and 60%) reduce concerns of external validity or sample selection bias.

[3] The organization does not use performance bonuses or other monetary incentives. The effect of introducing them

I generate exogenous variation in perceived gender shares in the job by showing the photograph of a real worker, who was randomized to be of the same or of different gender of the potential applicant.[4] While it is possible that photographs merely increase receivers' attention to the email or salience of gender, auxiliary surveys show that the two photographs induce an average difference in the perceived job female share of 6 percentage points (9% of the average female share). This treatment captures the fact that a predominantly female composition can affect men's choices by imposing a fixed cost on their utility, independently of occupational talent. For instance, working in a female-dominated job might threaten men's social image (Bursztyn and Jensen, 2017) and identity (Akerlof and Kranton, 2000, 2005). Men may also have an innate distaste for working with a majority of women (Becker, 1957) or anticipate employers' and customers' preferences for female workers.[5] These 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.

To shock expected returns to ability, I disclosed the aggregate performance of a selected past cohort of workers, which had either moderate or high success.[6] 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.[7] I interpret these statistics as signalling high and low marginal returns to ability on the job, respectively. Intuitively, lower past success (66%) signals that individual ability makes a larger difference in performance relative to a very high past success (89%).[8] Indeed, auxiliary surveys show that lower past performance makes high ability people increase their beliefs on the likelihood of being better than the median applicant from 38% to 45%, while low ability people reduce it from 37% to 32%. This second variation captures the fact that a predominantly female composition can affect men's choices indirectly, by imposing informational constraints, and interact with their job-specific talent. Men may not know and underestimate 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), but seeing only a few highly-selected members of their own gender creates uncertainty on the possibility to get rewards for talent in female-jobs (Arrow, 1998).[9]

I use a Roy-type framework to formalize how policies addressing these two channels affect the number and type of men who select into female-jobs. Candidates decide whether to apply for a female-

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.

[4] I draw on the design of audit (Bertrand and Mullainathan, 2004) and priming (Benjamin et al., 2010) studies.

[5] Some authors find evidence of 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).

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

[7] Being a high-performer means getting the highest assessment in practice tests when interacting with customers. See Section 3 for the exact wording.

[8] Notice that the reactions to the information manipulation would differ if potential minority applicants are trying to infer the likelihood that the employer will discriminate against them. Information indicating low performance could signal that the employer is statistically discriminating, which would generate a negative reaction by men.

[9] People might care about rewards to ability for extrinsic reasons, if performance is tied to incentives or career promotions, or intrinsic motivation, if they care about social recognition, feeling competent or about the actual impact generated in the job. Men should be particularly interested in returns to ability if norms that elect them as household breadwinners skew their choices towards jobs with steep careers (Bertrand et al., 2005).

dominated job or to choose an outside option. They care about monetary earnings, workplace gender shares and to what extent their ability impacts the employer's output. To capture the informational disadvantage of being the minority, I assume that men's priors on returns to ability in female-jobs are more uncertain and with lower mean than women's (Phelps, 1972; Arrow, 1973). An increase in the perceived share of own gender in the job shifts expected utility, while a change in expected returns to ability affects the steepness of utility with respect to job-specific talent. Tools that leverage the former channel can attract more men, but hires increase only if men are negatively sorted in the job. Tools that increase expected rewards to ability benefit high ability applicants, but might discourage low ability people if the job appears to be more difficult. Thus changing expected returns to ability may improve the quality of applicants when there is either positive or negative sorting in female-jobs. In either case, the joint change in application rates and quality of the applicants identifies whether men (and women) are negatively or positively sorted in social work.

I find that perceived gender composition do not affect men's application behavior. Men apply slightly more when they receive a female rather than a male photograph, but the effect is small and not statistically significant (3.2 percent). This 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. This is also consistent 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 of their same gender.[10]

Expected high 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 encourages men to apply more than being informed about outstanding past outcomes. This contrasts with most role model interventions, whose standard design provides participants with optimistic information of past success (Porter and Serra, 2017; Breda et al., 2018; Del Carpio and Guadalupe, 2018). Crucially, my paper shows that information of high past success can be interpreted as signal of low returns to ability rather than the unconditional probability of success, which might encourage only low-ability people to apply for the job.[11]

The magnitude of the effect of information is large, which is a particularly valuable outcome considering that the treatment is costless for the employer.[12] I further quantify the economic relevance of this treatment by estimating my theoretical model in a discrete-choice framework. At mean ability, the estimated effect of the experimental variation in returns to ability on applications is comparable to a 16.6% increase of the wage in the job (an increase in the hourly wage from 16.5 to 19.24 GBP). I also find that the difference in application rates between the two information treatments is larger among men who have been exposed to gender-segregated labour markets. This shows that new information is more valuable for people with limited experience in female-dominated jobs, a fact which is consistent with the hypothesis that they hold more uncertain and/or biased expectations.

[10]The positive coefficient of the female photograph on men's application also recalls evidence by Bertrand et al. (2010), who show that adding female photographs in adverts increase the demand of credit by both men and women.

[11]High average success coupled with low returns to ability provides insurance for low ability people against failure.

[12]The effect is over half the effect size of doubling wages found in Abebe et al. (2019) and increasing wages by 33% in Dal Bó et al. (2013), which show increased application rates of 18% and 26%, respectively. The effect is a quarter of that reported by Del Carpio and Guadalupe (2018), a difference which can be attributed to either higher application costs or more outside opportunities for participants in my setting.

Average male quality is higher in the treatment with expected higher returns to ability than lower returns to ability. Applicants in the former group are better in terms of observable characteristics such as cognitive skills, volunteering experience and achievement of high grades in university. They also receive more job offers (50%) and are equally likely to accept them compared to applicants with low expected returns to ability. Crucially, once on the job, men attracted by higher returns to ability show a quarter of a standard deviation higher performance vis-à-vis the low returns to ability treatment. Assuming no spillovers on inframarginal workers, the performance of those hires induced to apply by the treatment is two thirds of a standard deviation higher.[13] Interpreted through the lens of the model, these results show that men are negatively sorted in the job and that the marginal male applicant is facing an outside option which has steeper returns to ability than the average applicant. Increasing expected returns to ability in the job consequently improves the quality of the applicants.[14]

I conclude by checking for a trade-off between men's entry and women's exit. A common limitation of field experiments is that they are silent on general equilibrium effects. Nevertheless, showing a male photograph allows me to simulate a counterfactual world, in which men represent a higher (perceived) share in the job and see how women behave as a result. I find that there are 7.5% fewer women's applications in the male vis-à-vis female photograph treatment. This decrease in women's applications benefits the employer, however, as women who applied in this treatment arm received a higher offer rate and perform significantly better once on the job than women in the female photograph treatment. This suggests that a higher proportion of male workers in this job can improve female selection by discouraging the least talented women from applying or accepting the job. I also find that women are insensitive to information provision on average.

I rule out several competing explanations for my findings, such as social comparison or on-the-job dating opportunities, and different interpretations of the experimental manipulations by exploiting information on candidates' background and using auxiliary survey data.

Taken together, my results suggest that breaking informational barriers to men's entry in female-dominated jobs might increase gender diversity, as well as 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 in typically female tasks have been central in the US debate around the conversion of unemployed men into service jobs, as they have different policy implications.[15] 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 experimentation. If men had correct priors on rewards to ability in female-dominated jobs, my results suggest that low-cost

[13]Offer rates and performance on the job are available only for the subset of people who succeed in the hiring process and, for the latter, also accept the job. To attribute differences in these variables to the causal effect of the experiment on selection, the identifying assumption is that the treatment affects the composition of the pool of applicants and not their effort or the employer's screening criteria. This is guaranteed by the double-blind nature of the design. I also provide empirical evidence in Table 5.

[14]Better men with high opportunity costs might also have a lower likelihood of accepting and keeping the job (Abebe et al., 2019). This seems unlikely in my sample, where I see that the predicted hourly wage in the U.K. job market is skewed towards the left to the earnings distribution.

[15]See, for instance, this New York Times article (Miller, 2017).

organizational practices, such as recognition for good performance, may still attract a more diverse and qualified pool of applicants.

I contribute to three main streams of literature. First, personnel economics studies on the effect of posted wages or amenities on candidates' application and quality (Dal Bó et al., 2013; Marinescu and Wolthoff, 2016; Ashraf et al., 2019; Deserranno, 2019; Abebe et al., 2019). Both the methodology and the analysis of my paper draw on this work, but I further show that posted ads might address information frictions which prevent minorities from applying for jobs which are uncommon for their demographics. By showing the importance of expectations of non-monetary returns to ability, I contribute to studies that explore how subjective expectations of earnings drive educational and occupational choices (Nguyen, 2008; Jensen, 2010; Zafar, 2013; Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015, 2018) and models of job search which relax the assumption of complete information (Conlon et al., 2018). Both the motivation and experimental design of the paper are related to work on identity (Akerlof and Kranton, 2000, 2005; Akerlof, 2017; Bursztyn et al., 2018) and stereotypes (Steele, 1995; Stone et al., 1997; Hoff and Pandey, 2006; Bordalo et al., 2016). I contribute to these studies by comparing the impact of identity and (beliefs on) economic incentives on career choices outside of the laboratory, in a natural field setting. My work is also related to several experiments on competition and gender (Niederle and Vesterlund, 2007; Wozniak et al., 2010; Dreber et al., 2014; Reuben et al., 2017). Preferences for competitive environments are another way in which returns to ability may enter the individual decision problem. This implies that the implications of incorrect inferences about returns to ability across occupations might be amplified through their interaction with preferences for competition.[16]

2 Institutional context

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 to either final year students from a variety of disciplines or current workers across all industries.[17] Workers are assigned to teams allocated to Local Authorities across England and earn a stipend which is comparable to the average entry salary in social services (26k GBP), primary school teaching (24k GBP) and nursing (22k GBP) in the UK. 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. The program is a fast-track into the public sector with opportunities for faster career progression than standard routes into the profession. After the first two years, 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.[18]

This is an ideal setting to answer my research question for several reasons. First, women historically

[16] Along this lines, Reuben et al. (2017) show that attitudes by gender are correlated with different expectations of earnings across occupations.

[17] Professionals and students in the same field are not eligible. Eligible applicants should have a bachelor degree with 2:1 or higher and have obtained at least a C in Maths and English pre-university qualifications.

[18] The selective nature of the program weakens applicants' concerns regarding low social status that are typical of this industry. See, for instance, this article about the recruitment crisis in social care (Whittingham, 2018). This feature makes informational and psychological constraints more likely to have a first-order effect on selection.

represent more than 75% of social workers across the developed world, as shown in Figure 1 for the US. Most of the skills needed for the job are social in nature and commonly associated with women's comparative advantage (Ngai and Petrongolo, 2017). For instance, the website O*Net lists active listening, speaking, reading comprehension and social perceptiveness among the top skills needed for the job.[19] The stable gender ratio and the required skillset explain why stereotypes about social work as a "pink-collar job" have been persistent and widespread.[20] Men might lack information to estimate their own likelihood to succeed in the job and face social costs from peer pressure and gender norms.

Secondly, informational constraints are particularly relevant in my setting. In contrast to other female-dominated service jobs such as nursing or teaching, the average citizen has limited direct exposure to social work.[21] The organization also targets both men and women of any experience level, across disciplines and industries. This recruitment strategy implies that my sample features substantial heterogeneity in background exposure to social work and, consequently, variation in the information that people have about the occupation.

Third, in both the US and UK, social work is expected to grow in the next decades. 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.1, Bureau of Labor Statistics 2019).

Figure 2 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 assessment 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.

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 (also labeled "candidates" from hereon) 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 immediately after registration. The email contains their candidate number, which is necessary to access the application process, and some basic information about the hiring process.[22] I introduce exogenous variation in the content of the invitation-to-apply email along two dimensions: perceived gender shares and expected returns to ability.[23] The two experimental conditions were cross-randomized in a fully nested design, leading to a total of four

[19]For more information about O*Net, go to the website O*Net Online.

[20]See, for instance, this BBC news article by Hemmings (2018).

[21]According to the Department for Education, the number of social workers in England was 7% of the total number of teachers in state-funded schools in 2018.

[22]Respondents who do not meet the eligibility requirements receive a standard rejection email.

[23]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. However, the brevity of the form and the low application rate after registration (between 50% and 60%) reduces concerns of external validity or sample selection bias. Moreover, a minimum interest in the job makes this the relevant sample from a policy perspective.

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 part in a research study and prevents recruiters' assessment of candidates from being influenced by their treatment.[24] I discuss each experimental manipulation in the following paragraphs.

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.[25] I use my theoretical framework to show that these two interpretations are observationally equivalent and provide manipulation checks that are consistent with the former one.

This manipulation identifies the utility given by the workplace gender composition (or related attributes), assuming that photographs affect choices mainly through changing perceived gender proportions. Various confounders might threaten this identification strategy, including ethnicity: if white female candidates apply more after seeing an email portraying a white woman than a non-white man, we wouldn't 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 create negative emotions and anxiety in non-white subjects, as suggested by a rich literature on stereotype threat (Steele, 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. White people received pictures of white people and non-white people received pictures of non-white people (randomizing gender).[26]

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 didn't feature in other advertising campaigns or multimedia content from the organization for the duration of the intervention (until March 2018). This eliminates unobserved

[24] At registration, participants had to agree with the organization's data policy, which allowed for the possibility of impact evaluations and data sharing for evaluation and monitoring purposes. Anecdotally, participants thought that treatment emails were part of standard organizational practices.

[25] My main interpretation of the photographs 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).

[26] 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.

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.[27] 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 present the results of a complementary survey I conducted on Amazon Mechanical Turk to check for differences between people portrayed in the photographs, such as friendliness, attractiveness or work satisfaction.

Variation in expected returns to ability. This type of variation is difficult to induce for several reasons. In the ideal world, one would like to communicate to each person what their expected impact 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.[28] Moreover, the effect of individualized information on beliefs depends the level of people's priors, which was unobservable to me.

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 3):

> 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 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), thus these statistics refer to the social output obtained by previous workers. Both statistics were computed using actual records of the organization. This enabled to communicate truthful but partial information, which on average creates a wedge in beliefs between experimental groups (Dal Bó et al., 2017).[29]

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 percentage 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.[30] Updating on the returns to ability in the job, in turn, affects the expected performance for both low and high ability people and increases the differences between them. As a high ability person is more likely to perform well in a challenging job, her expected difference in

[27]The background portrays the real courtyard of one of the offices where workers are located.

[28]Generating a prediction based on observables was impossible for legal reasons, but also unlikely to reflect common practices.

[29]The need to communicate truthful information limited the range of possible statistics that I could use. The chosen ones were the most related to the constraint I am interested in studying and had the largest gap between cohorts.

[30]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.

impact should be greater between the two treatment groups than that of a low ability person. Low ability people might even be discouraged by a lower past success.[31]

This manipulation identifies the effect of expectations of returns to ability under the assumption that statistics of performance 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 10.

I reported information about on-the-job success in frontline interactions with clients for several reasons, primarily to induce variation in people's beliefs of their effectiveness in generating output for the employer.[32] 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).[33] Finally, the scores received in practice tasks are the joint outcome of workers' skills 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).[34]

Figure 3 shows an example of treatment email. From hereon, 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 information (which is 89%) while H refers to high returns to ability information (which is 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.[35]

Between November and December 2018, I administered an online survey to 565 people belonging to two distinct 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. Job applicants of the following year are very similar on observables and also capture possible unobservables that people interested in this particular job and organization share. I selected the sample on Prolific Academic by matching the composition of the field sample on several observables criteria.[36] Both samples were incentivized for participation and the survey had an average completion

[31] The given interpretation of the information manipulation relies on the assumption that experimental subjects keep fixed the range of abilities that workers in the job have in both treatment groups. Anecdotally, this is consistent with the high reputation that the organization has as a selective employer.

[32] Information on the probability of getting a job offer was also available, but it would have been less appropriate for my research question and 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, 1995), also on white people (Stone et al., 1997). By being long-term outcomes, the chosen statistics can affect beliefs about expected returns on the job, while avoiding negative emotional reactions with direct implications on short-term performance.

[33] 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.

[34] Men and non-white people should be less likely to apply when seeing that the job is more difficult if they fear discrimination by the clients. The opposite effect would thus exclude this interpretation.

[35] I couldn't directly elicit participants' beliefs on gender proportions and expected returns to ability because the survey could have interacted with reaction to the treatment (for instance, by making gender too salient).

[36] I selected participants on Prolific Academic to match the share of people in full time employment, who studied

time of 15 minutes. Appendix B describes the sampling strategy and questions in detail.

In a between-subject design, I randomly assigned respondents to see one of the four treatment emails used in the field experiment. After mandatory understanding checks, the survey elicited beliefs on a variety of characteristics of the job and the pool of applicants. Figure 4 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. 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 by clients).

Testing whether people update expected returns to ability in the job requires two ingredients: knowing their approximate position across the ability distribution and their corresponding returns. The left panel of Figure 5 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.[37] The right panel of Figure 5 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 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 the 89% statistic, as demonstrated by the larger difference in expected rankings between high and low ability people in the former case.

The identification of the effect of perceived gender shares separately from expectations of returns

subjects related to social jobs and of non-white ethnicity in my field sample. All people are from the UK and of age between 18 and 64.

[37] 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 being 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.

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 employer's messages affect expectations of returns to ability ("expectations effect") and utility from gender composition on the job ("gender effect"). The main goal is to guide the empirical analysis and generate predictions on the size and quality of the applicants' pool in each treatment group and for different parameters' ranges.

4.1 Environment, preferences and beliefs

Potential applicants are characterized by group belonging g and ability a_i. Everyone can observe 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 $a_i \sim U[.]$.[38] They 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: $U^o(a_i) = w_g^o + v_g a_i$. 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:

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

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. I define the difference between the wage in the job and in the outside option (both known) as $\overline{w}_g = w_g^o - w$.[39]

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), need for feeling competent (Elliot and Dweck, 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. [40] In this view, θ_g can be

[38] 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. The literature on overconfidence shows that a reversal in rankings is atypical (see Moore and Healy, 2008, Coffman et al., 2019).

[39] 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.

[40] In 2016, the partner organization asked 83 applicants about their motivations for applying. 51% mention career opportunities, 37% mention social impact and 31% mention the "challenge" of making things better in local communities.

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, while ability levels above (below) $\hat{a}_g$ get higher utility from higher (lower) θ_g.[41]

The component $\alpha_i s_g$ formalizes agents' utility from workplace gender composition, which I assume to be linear in the share of their own gender g. 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 $s_M = 1 - s_W$. I assume that $\alpha_i \in [0, 1]$, meaning that people prefer working with 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, social image concerns (Bursztyn and Jensen, 2017) and threats to identity (Akerlof and Kranton, 2000, 2005) might be particularly important for men.[42]

The second source of uncertainty is in returns to ability. Agents don't perfectly know how much reward they are going to get from being above the minimum ability requirement. A key feature of this framework is that 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: $\overline{\sigma}^2_M \geq \overline{\sigma}^2_W$.

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.[43] Appendix F.1 provides empirical evidence that men tend to have lower and more dispersed expectations of own group's performance in social work than women.[44] 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 application rates from one or both gender groups.[45] The employer includes two pieces of information: the photograph

[41] In Lazear et a. (2018), this is the marginal worker whose productivity is the same in hard and easy tasks.

[42] 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 (McPherson et al., 2001; Jackson, 2009). The work by Akerlof and Kranton (2000; 2005) assumes that choosing an activity which is uncommon for 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.

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

[44] This evidence come from the auxiliary online experiments described in Appendix B.

[45] 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)

of a worker, who can be a man or a woman, and information on the difficulty of the job. As the employer's profits are increasing in the quality of the workforce, the information provided aims at increasing not only applications' numbers, but also the quality of applicants.

Recruitment messages are a vector (P, S) such that $p \in \{M, W\}$ 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 hereon, I will denote the **experimental realizations** of the signal $s \in \{s_L, s_H\}$.[46] I will maintain this definition of s_L and s_H throughout this section.

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. The following paragraphs describe the updating process in period 1 in detail.[47]

Pictures $p \in \{M, W\}$ contained in the posted advertisements have a direct utility effect by changing perceived gender shares. Seeing a photograph of gender g will increase the perceived share of that gender in the job: $E[s_g|p = g] > E[s_g|p \neq g]$. If the predominantly-female composition discourages men from applying, seeing a person of the same gender will increase men's expected utility on the job. An alternative way in which photographs can affect utility is through α_i, by changing the salience of gender (as in priming studies, for instance, Benjamin et al., 2010). I do not disentangle these two explanations, but I showed manipulation checks consistent with a change in s_g (Section 3.1).

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 information treatments:

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

or organizational reputation (Erhardt et al., 2003; Carter et al., 2003; Flabbi et al., 2019).

[46] 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.

[47] I assume that advertisements do not affect the knowledge of individual ability a_i. This is a common assumption in the literature (Ashraf et al., 2019; Abebe et al., 2019). It is also consistent with evidence described in the manipulation checks and the fact that information was about aggregate performance and not about people similar to the recipient. Another way of capturing the crucial message of the model if allowing a_i to be updated as an effect of the intervention, but increasingly in a_i. This can happen, for instance, if people have more precise prior beliefs about their own ability in the left than right tail of the distribution, or if people are underconfident at the higher end of the ability distribution.

$\Delta\theta_g$ 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 $U^j(a_g^*) = U^o(a_g^*)$.[48] 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 $U^{j'}(a_i) = \overline{\theta}_g$ and $U^{o'}(a_i) = v_g$, 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 $U^{j'}(a_i) < U^{o'}(a_i)$, 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 $U^{j'}(a_i) < U^{o'}(a_i)$: 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. Figure 6 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 top panel shows the case illustrated by result 1 ($U^{j'}(a_i) < U^{o'}(a_i)$) and the bottom panel shows the alternative case ($U^{j'}(a_i) > U^{o'}(a_i)$).

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 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.[49] Define $B = \alpha_i s_g - \overline{w}_g - c - v_g\hat{a}$.

[48] See Appendix F for the formal proof.

[49] 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 $\overline{\theta}_g - v_g$.

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.[50] Figure 7 shows the graphical intuition for Result 2. The top panel shows the case illustrated by Result 2 ($U^{j'}(a_i) < U^{o'}(a_i)$) and the bottom panel shows the alternative case ($U^{j'}(a_i) > U^{o'}(a_i)$). 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 utility on the job.[51] 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. The dashed red line shows this in the top panel of Figure 7.

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. Changes in expected returns to ability benefit high ability applicants, but might discourage low ability people if the job appears to be more difficult. This implies that changing expected returns to ability can potentially improve the quality of applicants when there is either positive or negative sorting in female-jobs.

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.[52] The proportion of people from lower socio-economic backgrounds is substantial: 19% of subjects come from families where parents had an unskilled occupation, 27% of subjects received economic support in school and 2% were looked after by a social worker as a child.[53]

[50] If $B < 0$ and $U^{j'}(a_i) < U^{o'}(a_i)$, when $s = s_H$, N_g is smaller and both the marginal and average abilities are lower than when $s = s_L$.

[51] 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 $\hat{a} = 0$).

[52] As the student population in these universities represent 15% of higher education institutions in the UK, the program disproportionally attracts students coming from selective universities. I define top-tier universities as those belonging to the Russell group. The Russell Group Profile 2017 is available here

[53] IA care leaver is a person who has been looked after by a local authority for at least 13 weeks since the age of 14. n 2012, the total number of care leavers represented 0.12% of the total UK population between 16 and 25, while they 1.6% of applicants up to 25 years old. Estimates are based on the 2011 UK Population Census (available here) and the 2012 "Care leavers in England data pack" by the Department for Education (available here).

Almost half of the sample (41%) currently work full time (FTE from hereon), mostly in the public sector or healthcare, but a substantial share also comes from science, business or technology.

Men and women tend to have a similar socio-economic background and experience with the organization, but differ in 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.

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.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).

Table D.1 compares the experimental sample with a random subsample from the UK Labour Force Survey (LFS) with the same age distribution. 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 and more likely to have worked in the public sector or healthcare. These differences confirm that people in the 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. For instance, men in the sample might care less about gender composition 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. Secondly, participants to the experiment might have different outside options than average LFS respondents (differing in parameters such a v_g or w_g^o). This implies that selection on talent could be different in other samples facing different structural parameters. Nevertheless, I think that there is scope for generalizability as this is a relevant sample for policy. Conditional on interest in the sector, the experimental pool is representative of job applicants to similar programs.[54]

5.1 Main specifications and identification assumptions

In the following sections I present evidence on the effect of photographs and information on the applicant pool's size, quality and performance on the job. The empirical strategy relies on the independent random assignment of these two manipulations.[55] I perform separate estimations for men and women. Given the nature of the job, the marginal female and male applicant might be very different from each other, thus a fully interacted model seems the appropriate specification.

[54] For instance, people in my sample resemble applicants for Teach For America (Coffman et al., 2014).

[55] The partner organization was interested in which of the recruitment messages worked best in increasing applications compared to their standard email. This comparison 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 (Gonzales and Loureiro, 2014; Bertrand et al., 2010). I thus only compare treatment emails with each other, leaving aside the pure control email.

Consider a potential applicant i that decides whether to apply ($y_i = 1$) or not. My main specification is the following:

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

where Pic_i^M is a dummy equal to one if i was assigned to receive a male photograph and $Returns_i^H$ is a dummy for the high returns to ability information. The vector of controls X_i contains the following variables: dummy for non-white ethnicity, 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.[56]

I define the application dummy y_i to be one if a candidate submits the application form and keeps showing up at any later stage of the selection process, conditional on reaching that stage. This definition implies that I do not consider as "applicant" a candidate who is called to the second round of interviews but decides to not show up. This variable thus represents the cumulative effect of the treatment throughout selection stages.[57] Other outcomes will be whether i receives a job offer (conditional on application), where she/he accepts and average performance scores on the job. 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 following is independent of treatment assignment. This was guaranteed by the double-blind design of the experiment (see Section 3).

In model (2), coefficient β_1 tests the null hypothesis of no effect of perceived gender shares on applications. Failing to reject the null indicates that either the treatment does not change perceived gender shares ($ds_g = 0$) or that the workplace gender composition does not affect application decisions ($\alpha_i = 0$). Coefficient β_2 tests the null hypothesis of no effect of expected returns to ability on applications. Failing to reject the null indicates that either people do not update their expected returns to ability ($\Delta\theta_g = 0$) or that the ability of the marginal applicant is so close to $\hat{a}_g$ that changes in marginal returns to ability do not affect utility. Parameters 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. In order to check whether this assumption is empirically valid, I can combine the two manipulation to study whether there is an effect of their interaction.[58] I use the following specification:

$$y_{ig} = c + \delta_1 \cdot Pic_i^g \cdot Returns_i^H + \delta_2 \cdot Pic_i^{-g} \cdot Returns_i^L + \delta_3 \cdot Pic_i^{-g} \cdot Returns_i^H + X_i'\lambda + \epsilon_i \tag{3}$$

where Pic_i^g (Pic_i^{-g}) is a dummy equal to one if i was assigned to receive a picture of the same (opposite) gender and $Returns_i^H$ ($Returns_i^L$) is a dummy for high (low) returns to ability information. Specification (3) uses the email that combines the same-gender picture with low-returns information as omitted category.[59] Model (3) tests three null hypotheses: $\delta_j = 0$, with $j \in \{1, 2, 3\}$.

[56] Results are robust to adding a bias-reduction modification, which is analogous to the modification by McCaffrey and Bell (2002), as proposed in Imbens and Kolesar (2016).

[57] In order to apply for the job, candidates have to submit an application form and take an online test within seven days of the application submission, for an estimated time of completion between 4 and 6 hours. The application form contains motivational questions and several sections on qualifications and employment experience. The average application rate across years is 60% of registered candidates and it is higher for women than men (by 5 to 10 pp).

[58] This model has to be taken with a grain of salt as the study is underpowered to look at the interaction.

[59] This control group seems also a natural benchmark inspired by many studies in psychology and economics which

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; Gerber and Green, 2012).[60] 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: men's entry

6.1 The effect of a shock to perceived gender shares

A higher perceived share of own gender in the job does not affect men's applications.[61] Receiving an email with a male person reduces men's applications by 1.8 percentage points with respect to an email featuring a female person (Column (1) of Table 2). However, this coefficient is imprecisely estimated and I cannot reject the null hypothesis of no difference between the two photographs. This is a surprising null result in light of many policy proposals that try to attract men in female sectors through ads portraying people of the same gender (Abadie, 2018).[62]

One way to reconcile this evidence with current policies is thinking about sample selection. If self-selection into registration is negatively correlated with tastes for workplace gender composition (or tastes for correlated attributes), men in my sample could potentially have a lower α_i than the average man. This implies that the estimated effect of perceived male shares is a lower bound of what should be expected for the average man. Nevertheless, in a complementary experiment with the same organization, I show that ads portraying men are not enough to encourage even a more common population of male students to apply (see Appendix H). This extends the external validity of the null result of Table 2 and implies no role for that either gender homophily with co-workers or social stigma in men's choices.

This 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. Moreover, data from the US between 1970 and 2018 show that the wage gender gap is smaller in female-dominated than in male-dominated occupations.[63] This evidence goes against the hypothesis that men might get compensating differentials for a distaste in predominantly-female occupations. This null result is also consistent 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. The positive coefficient of the female photograph on men's application also relates to Bertrand et al. (2010), who show that female photographs increase the demand of credit

attempt to increase minorities' performance expectations through successful role models (Dasgupta and Asgari, 2004; Cheryan et al., 2011). Moreover, qualitative interviews and focus groups indicated that priors are relatively closer to the "Low Returns" information.

[60] I use the code provided by Alwyn Young on his website.

[61] Figure 8a shows application rates across treatment groups from the raw data.

[62] See, for instance, the article "Male Nurses: not just a woman's job" in The Economist (August 2018) at this link.

[63] Data are from the CPS March supplement. Tables are not reported in this paper.

by both men and women through a non-deliberative reaction to adverts' creative contents. However, it's not clear whether such a System-1 effect should arise in my setting, where photographs represent testimonials of previous applicants and aim to trigger the recipient's comparison between them.[64]

6.2 The effect of a shock to expectations of returns to ability

Men react strongly to the expectations manipulation. This is shown in the bottom row of Column (1) of 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 the omitted category, with a p-value of 0.04. This represents 14% of the mean in the low expected returns treatment and 12% of the pure control mean.

In other words, men's entry into this job is positively affected by information of lower past success among workers. This result is novel and contrasts many role model interventions, whose standard design provides high statistics of success to minority members to increase their perceived likelihood to succeed in uncommon jobs. For instance, Del Carpio and Guadalupe (2018) show that girls are more likely to apply for a coding boot camp if they are first exposed to information on a same-gender role model, availability of female networks and high probability of success in the tech sector. 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 might contribute to explain why, on average, Del Carpio and Guadalupe get negative selection in their experiment.

The increase in application rates in the high expected returns to ability treatment suggests that the marginal applicant has ability a^* greater than pivot ability $\hat{a}$. In the opposite case, higher expected returns to ability could even attract less applicants. The theoretical interpretation of the treatment as a rotation of expected utility on the job with respect to ability also implies that the change in application rates should be positively correlated with job-specific ability. I test this in Table A.1, where I show that the effect of higher expected returns to ability on applications is stronger among men with above-median predicted performance on the job (Columns (1) and (2)) and linearly increasing in this proxy of job-specific ability (Column (3)).[65]

This result suggests that informational constraints might be important barriers to men's entry in female-dominated jobs. It is surprising that limited information plays a role in my context, where one could assume there are nearly unlimited opportunities for learning and experimentation. But 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 even simply pay attention to - information on careers that are uncommon for their gender. This might be especially the case for people with a more valuable outside option.

[64] I discuss this point further in Section 10.

[65] I compute predicted performance on the job using baseline variables that are available for everyone. I use the observed scores on the job to impute the predicted score to an individual with missing actual score using a linear truncated regression. I use the following set of 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. The implicit assumption is that the way in which these variables affect on the job performance is independent of being hired and treatment status. See details in Section 9.

Men's reaction to the information treatment also uncovers the importance of expectations of non-monetary returns to ability in their choices. A rich literature in labour economics explores how subjective expectations of earnings drive educational and occupational choices (Nguyen, 2008; Jensen, 2010; Zafar, 2013; Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015, 2018). My information treatment does not change expectations of incentive schemes or earnings, but more broadly the extent to which people think their talent will be rewarded in the job. This seems an important dimension to complement the traditional view of returns to ability. My result is also in line with recent models of job search that find that the costs of information frictions are sizable, but can be mitigated by learning (Conlon et al., 2018).

How large is the increase in men's applications in the most successful treatment? Applications increase by 12% in treatment (W,H) as compared to the pure control group. This effect is a quarter of the one reported in Del Carpio and Guadalupe (2018), but this gap can be explained by differences in the application/opportunity costs between settings as well as the level of application rates in the control group (7% in their setting versus 53% in my context).[66] However, the magnitude seems relatively large when compared to the effect of increasing posted wages in Abebe et al. (2019), who get an increase in application rates of 18%, and Dal Bó et al. (2013), who show an increase of 26%. An effect between two thirds and a half of the one obtained in these papers is quite substantial given the light touch nature of my treatments, which were nearly costless to the employer.

Appendix Table A.2 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. In other words, in men's aggregate sample, I cannot reject the null hypothesis of equal effects of emails (W,H) and (M,H), on the one hand, and (W,L) and (M,L), on the other. This means that the additivity assumption used in the standard version of the model seems appropriate.[67]

6.3 From the model to the data: heterogeneity by gender norms and priors' uncertainty

Do gender composition or expectations of returns to ability matter relatively more for men not used to seeing other men in the job? The model comparative statics predict 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 teenagehood* 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 (Blau et al., 1998; Akerlof and Kranton, 2000, 2005; Goldin, 2014; Cortes and Pan, 2018).[68] Exposure to gender segregation can also affect the persistence of biased beliefs on group ability, an insight used by Arrow (1973, 1998) to explain the

[66]In my experiment, completing the application form takes between 4 and 6 hours, almost ten times more than in the Guadalupe and Del Carpio's setting.

[67]Manipulation checks discussed in Section 3.1 also gave reassurance of this assumption.

[68]While gender norms may cause men and women to choose different occupations, exposure to job gender segregation may in turn make people internalize gender norms which make occupational differences persist over time (Blau and Kahn, 2000, 2017; Charles et al., 2018; Bell et al., 2019).

persistence of long-term statistical discrimination.[69] I posit that a similar channel can limit minorities' knowledge of their own returns to ability in uncommon jobs.

The construction of 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, 1955), which identifies the percentage of women (or men) that would have to change occupations for the occupational distribution of the two genders to be equal.[70] Using a bridge, I merged the index with my experimental data through the subjects' secondary school postcode and, when missing, home postcode.[71] I use this index as an individual level measure of exposure to gender-segregated labour markets in the decade prior to the job application, under the assumption that the choice of residence is not affected by the index itself.

Table 3 estimates heterogeneous treatment effects 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 Table 3 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.

The main caveat for interpretation of Table 3 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.[72] Columns (3) and (4) of Table 3 repeat the same exercise using a different index: the average share of men working in female-dominated occupations in the local labour market.

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 U.K. 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.

[69] 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".

[70] 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 of the local labour market. It takes values between 0 (complete integration) and 1 (complete segregation).

[71] 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 3 using only the subset of people with data on school location, but power drops.

[72] 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. Table C.1 compares men and women coming from areas with high versus low gender segregation on a variety of observables.

6.4 From the model to the data: heterogeneity by outside option parameters

Increasing the size of the applicants' pool is desirable for the employer as long as it allows hiring of better workers.[73]. In the model, this depends on whether potential applicants are facing steeper returns to ability in the outside option or in the job. In this section, I show that men with steep returns to ability in the outside option (high v_g) are more likely to apply in the high expected returns treatment than men with a flatter outside option. This suggests that the information treatment might not be generating a quantity-quality trade-off for the employer.

In Table 4, I split the sample of candidates by above/below median 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.[74] Columns 1) and 2) of Table 4 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 is consistent with the theoretical case $\theta_M < v_M$ and suggests that we should expect the marginal applicant in the high expected returns treatment to be better than in the low expected returns treatment.

The theoretical model further indicates that the effect of expectations of higher returns to ability on application likelihood depends on the marginal applicant's position in the ability distribution. Higher expected returns are predicted to attract more applications than lower expected returns for high ability people. However, the positive difference is predicted to be decreasing in a_i and to become negative as a_i becomes lower than $\hat{a}$. Heterogeneous treatment effects with respect to the outside option level w^o provide evidence of this. The last three Columns of Table 4 repeat specification (2) splitting the candidates' sample by quantiles of individual outside option. An individual's outside option is their expected hourly-wage in the U.K. 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. Appendix D contains a detailed summary of the methodology used.

The evidence reported in Columns (3), (4) and (5) of Table 4 is consistent with the information treatment being a change in slope of expected returns on the job. 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 is what we should expect if men's sorting in the job is negative: the difference in slopes $\Delta\theta_g$ implies a bigger difference in application rates among low percentiles of the outside option, where the marginal applicant has a relatively higher ability level (top panel of Figure 7).

Overall, this section suggests that men are negatively sorted in the job and, consequently, that the larger pool of applicants attracted by raising expectations of returns to ability should also be more

[73]Net of the additional screening costs

[74]This index of wage dispersion is a function of the endogenous choice of university subject made by the candidates. Thus Table 4 could capture heterogeneous treatment effects due to other unobservable differences between candidates who chose the same university subject. As a robustness check, Table A.5 repeats the exercise using the wage dispersion of the region where each candidate lives, under the assumption of limited mobility across regional labour markets. In the LFS, only 16% of workers work in a region which is different than their region of residence (excluding people working in Central London and commute).

qualified. I provide evidence consistent with this prediction in the next section.

7 Results: men's quality

In this section I show two main results on applicants' quality. First, male applicants in the "High Expected Returns" treatment are better on observables and receive a higher offer rate than in the "Low Expected Returns" treatment. Secondly, once on the job, male workers from the "High Expected Returns" treatment have better qualifications, perform better and state that are more likely to stay in social work than in the "Low Expected Returns" treatment.

7.1 Applicants' skills and job offers

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 correlated with receiving a job offer. I construct an index which averages the following (standardized) 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.[75] Appendix Figure A.2a shows 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. The positive gap between the two treatments is positive across the distribution, but slightly higher in middle quantiles (Table E.1). Men in the male photograph treatment are also better in the same observables on average, but the effect is driven by the highest quantile of the distribution (Table E.1). This is consistent with an improvement at the margin generated by the higher expected returns to ability information and the male photograph.

Men in the high expected returns treatment consequently get more job offers than applicants in the low expected returns treatment. The offer rate is 18%, which is 6.2 percentage points higher than in the treatment providing information of low returns to ability. This is shown in Column (2) of Table 2, where the dependent variable equals one if a person received a job offer, conditional on applying.

To attribute the increased offer rate to the causal effect of the treatment on applicants' composition one needs to exclude that the treatment affects the employer's screening criteria (Ashraf et al., 2019). I check this in Table 5, which shows the coefficients of the following regression:

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

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 desirable qualification. In addition to the set of variables

[75] To define cognitive and manual 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. For each person, the measures of cognitive and manual skills should be interpreted as the average level of cognitive and manual skills acquired in past work experiences.

defined above, I consider also having studied a subject aligned with the content of the job and having received the maximum score in Maths pre-university qualifications.[76]

Columns (1) and (3) report the coefficients $\alpha_j^{T^1}$ and $\alpha_j^{T^2}$ for the information and photograph conditions, respectively. First, they show that 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 doesn't seem to matter. Columns (2) and (4) report the p-value of tests of equality of coefficients $\alpha_j^{T^1} = \alpha_j^{T^2}$. Most of the reported p-values are above 0.20, indicating that I cannot reject the null hypothesis of equality of the employer's selection criteria across treatments. Two comparisons out of twelve 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.[77]

7.2 Workers' skills and performance on the job

In this section, I show that raising expectations of returns to ability allows the employer to select male workers that are better in terms of observable qualifications and perform better on the job. For this exercise, I consider the subset of job offerees who accepted the offer (62 out of 88 men, of whom 43 in the treatment groups). They all start working for the organization in July 2018. After a first month of training, they are sent to their allocated team across communities in different UK regions.[78]

Figure 9a shows the difference in the proportion of men who hold a certain qualification between treatment groups, using the same set of variables defined in the previous section. The Figure on the left shows differences between the male and female photograph and the Figure on the right shows differences between high and low expected returns to ability. Men are better in terms of observables in both the male photograph treatment (vis-à-vis the female photograph treatment) and in the high expected returns to ability treatment (vis-à-vis low expected returns to ability). For instance, 38% of men in the high expected returns to ability treatment and 23% in the low expected returns to ability treatment achieved a first grade in university. This shows that the employer is able to hire better workers through the improvement in the quality of the pool of applicants generated by the treatment.[79] However, a limitation of this evidence is the small sample size. I then turn to performance on the job, where the availability of repeated measures for each person reduces concerns on sample size.

[76] I define "aligned subjects" 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 consider aligned subjects also those titles that combine different disciplines, such as "Law and Economics".

[77] I also ran the same specification adding measures of cognitive and manual skills inferred by the employment history reported by the candidates. Results are robust to this inclusion. The employer selects people with higher cognitive skills, but manual skills are deemed less important. There are also no differences in the extent to which cognitive and manual skills affect the probability of receiving an offer across treatments.

[78] 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.

[79] I show dynamics of observable qualities over the hiring, stage after stage, in Appendix G.

Measuring performance in public sector frontline jobs is rare. A convenient feature of my partner organization is that workers are continuously assessed in both theoretical and practice tests. I measure workers' performance using the grade they received during the first six months on the job, which is the period covered by the data available so far.[80] Between August 2018 and January 2019, new workers are evaluated in five different assessments: a first-month performance review, three theory assignments (e.g., case studies, essays) and one practice evaluation.[81]

Figure 9b shows the distribution of men's average test scores by experimental treatment. The distribution of test scores is shifted to the right for men in the high expected returns vis-à-vis the low expected returns to ability treatment, with a bigger difference at the left tail of the distribution (right-hand side figure). This suggests that higher returns to ability improve the male workforce through a change in composition which attracts better men and, at the same time, deters the worst men from entering the job. This corresponds to the theoretical case in which the posterior beliefs on θ_g are higher than returns to ability in the outside option. The model doesn't predict the right shift in the distribution of scores for men in the female photograph vis-à-vis the male photograph treatment (left-hand side figure).[82] However, 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.

I estimate the following model using panel data at the worker-assessment level:

$$score_{ia} = \alpha + \beta_1 Pic_i^M + \beta_2 Returns_i^H + X_i'\lambda + \epsilon_{ia} \tag{4}$$

where $score_{ia}$ is worker's i grade in assessment a normalized by the mean and standard deviation of male workers' grades.[83] The vector X_i includes, in addition to the basic controls of specification (2), dummies for the region where the worker is allocated, a dummy for whether the worker has been allocated to his preferred region, a dummy for whether the worker studied in a top tier UK university and the score she got in Maths pre-university qualifications (as proxies for baseline ability). Standard errors are clustered at the worker level.

Parameter β_1 and β_2 measure the causal effect of the experimental manipulations under the identifying assumption that the change in observed job performance is due to a change in applicants' selection caused by the treatment. Another identifying assumptions is of no contamination across treatments, which seems reasonable in this context given the time lag between hiring and working.

The bottom row of Table 6 shows that men with high expected returns to ability perform significantly better: their scores are 24% of a standard deviation higher than men with low expected returns (p-val < 0.10). The effect increases (to 0.36) when accounting for the fact that men in the high expected returns to ability treatment tend to be allocated to more challenging communities. This is

[80] I will keep tracking participants for the full duration of the programme until July 2021.

[81] The theory assessments 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 "potential" of each worker of doing a good job once in the local communities. 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 assignment or even that an experiment took place.

[82] This difference is driven by non-white men: they are concentrated in the male photograph treatment and they all perform significantly and substantially worse than others (in the order of 20% lower scores on average).

[83] In the raw data, each grade is on a scale between 0 and 100, where 40% is the minimum threshold for passing and grades above 70% correspond to a distinction.

shown in Column (2) of Table 6, which adds weights for the difficulty of the local community where a worker is allocated to.[84] Appendix E uses quantile regressions to show that the improvement in men's test scores in the high expectations treatment is concentrated among the lowest quantiles.[85]

Table 7 shows additional results on the attitudes and perceptions of men hired in the job. The main result is that men in the high expected returns to ability and in the male photograph treatments are more likely to say that they would like to stay in the job in the future. This is important in a sector where as many as 50% of workers plan to stay less than two years (Ravalier, 2018). Future data collection will shed light on whether these intentions turn into higher retention.

7.3 Selection or self-fulfilling prophecy?

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 such an effect is a self-fulfilling prophecy: believing in higher chances to be successful might make men put more effort and motivation over the hiring process, with a subsequent higher offer rate (but not necessarily a change in selection). Such an effect has been documented in a few papers as a response to varied leaders' expectations (Rosenthal, 1994; Eden, 1992; Eden and Ravid, 1982) or to prejudice against minorities (Benyishay, 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.6 shows that men in the two information treatment groups do not differ in the effort put in application completion, as measured by the percentage of fields filled-in and the number of characters used to answer the application questions. Secondly, workers in the high expected returns treatment are better, on observables, than workers with lower expected returns. 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 two implications of this hypothesis: performance effects should be waning over time and be driven by people surrounded by worse colleagues. Figure A.3 shows that there is no decreasing trend in the coefficients on the treatment indicator variable in separate regressions for each of the five on-the-job assessments. I also don't find evidence of a greater performance by men in teams with a lower leave-out-mean in the high versus low expected returns treatment.[86]

[84] 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). The report includes data on the outcomes of inspections carried out on all registered social work providers in England. For more information, consult the Department for Education website here and Ofsted website here. There are 52 local authorities where workers in my sample are 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 Ofsted's scores on helping children, child care, leadership effectiveness.

[85] The differential change at different points of the ability distribution is possible because the employer makes job offers on a rolling basis and ranks all the candidates independently of treatment assignment in a centralized way.

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

8 Will men's entry into female-dominated jobs affect women?

Encouraging men's entry in female-dominated jobs inevitably affects women, so the net benefit for the employer is unclear if we ignore the effect that increasing minorities' participation will have on the majority in the job. If male shares in female-dominated jobs actually increased, would there be any negative impact on the number and quality of female applications? I use the photograph manipulation to answer this question. Showing a male photograph allows me to simulate a counterfactual world in which people perceive the share of men in the job to be higher and see how women behave as a result.

I find that a higher (perceived) male share discourages women from applying for the job.[87] Column (1) of Table 8 shows that there are 7.5% fewer women's applications in the male vis-à-vis female photograph treatment. An alternative way of interpreting this result is that women infer that the employer is looking for men, but this doesn't rule out that the effect is driven by an anticipated future change in gender composition. This alternative story would still indicate that employer's active policies to attract more men in female-dominated jobs might discourage women from applying.

Fewer applications by women turn out to be a positive outcome for the employer. Women who applied despite seeing a male photograph receive a slightly higher offer rate (not statistically significant), are more likely to accept the job and perform significantly better in the workplace than women in the female photograph treatment. This is shown in Columns (2) to (4) of Table 8. Column (4) shows that women in the high perceived male share treatment achieve average test scores on the job which are 19% of a standard deviation higher than women in the low perceived male share treatment.

Consistent with such increased performance, Figure 10 shows that female workers in the male photograph treatment are better on several observables characteristics (Panel A) and that the distribution of their on-the-job test scores is shifted to the right of the one in the female photograph treatment (Panel B). Appendix Tables A.7 and E.2 confirm these results exploiting repeated assessments for each person and introducing individual-level clustered standard errors to account for within-person correlations in the errors. Improvements in women's performance in the male photograph treatment are concentrated among the middle quantiles (between the $25t^{th}$ and $75t^{th}$, see Table E.2).[88]

The joint change in applications and quality suggests that, in contrast to men, the sorting of women in the job is positive. This is implied by the fact that fewer applications are correlated with an increase in average quality. Table A.4 confirms this conjecture by showing heterogeneous treatment effects by the degree of wage dispersion faced in the UK labour market on women's applications. As expected, the negative effect of the male photograph is concentrated among women with flatter outside options. This is consistent with the theoretical case $\theta_W > v_W$, in which the marginal applicant in the female photograph treatment is worse than in the male photograph treatment.

Women are insensitive to information provision on average, which is consistent with the majority holding more precise priors on their performance in the occupation. The second row of Table 8 shows that the point estimate on the $Returns_i^H$ dummy is -0.015 and far from being statistically significant. Column (1) of Table A.3 confirms that the two genders react differently to the expected returns

[87] Figure 8b shows application rates across treatment groups from the raw data.

[88] As in men's case, I can exclude that women's performance differences come from a "surprise" once people compare expected and actual gender shares on the job. Figure A.3 shows that the effect of the male photograph on female performance does not show a decreasing trend over time.

treatment. Overall, these results suggest that a higher proportion of male workers in this job can improve female selection by discouraging the least talented women from applying and/or accepting the job.

8.1 Do women care about workplace gender composition?

Women's reaction to the photograph manipulation could suggest that women value co-workers gender more than men do. This has been suggested by some authors (Haile, 2012; Lordan and Pischke, 2016), who show that women's well-being is higher in workplaces with a higher female share.

Yet, some evidence invites a more cautious interpretation of women's behaviour. First, I cannot reject the null hypothesis that men and women react in the same way to the photograph manipulation (Table 8). Secondly, the interaction between the two manipulations is important. The dashed lines of the right panel of Figure 11 show that women are less likely to apply when seeing a male photograph in combination with information of high returns to ability. When expected returns to ability are low, the impact of the photograph manipulation on women's application rates is reduced. This suggests that women's behaviour does not stem from a generalized preference for working with own gender.

There are alternative ways to interpret women's stronger reaction to a high perceived male share when in combination with expected high returns to ability. One hypothesis goes through preferences: women dislike working with men in more challenging environments, as suggested by the literature on gender differences in preferences for competition (Niederle and Vesterlund, 2007; Niederle and Yestrumskas, 2008). In the model, this explanation could be accounted for by making the individual utility weight on gender composition α_i a positive function of the returns to ability in the job.[89].

An alternative hypothesis goes through beliefs: gender shares affect women's inference of their expected returns to ability on the job. This is in line with work on beliefs about gender (Coffman et al., 2019; Bordalo et al., 2019), overconfidence (Croson and Gneezy, 2009) and a few results in the competitiveness literature (Wozniak et al., 2010; Dreber et al., 2014). If women have a comparative advantage in female-jobs, a lower female share might signal a decrease in such advantage, which becomes relatively more important in a job where ability matters more. An extension of my model that allows gender shares to impact expected returns to ability might account for this mechanism. I present such extension in Section F.3, where I assume that gender shares impact expected on-the-job ability by providing information on the pivot ability level $\hat{a}$. 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.

9 Estimating structural parameters

One limitation of the experimental design is that I don't observe people's updating of their expected returns to ability or the utility weight they give to the workplace gender composition. In this section, I estimate these parameters using a discrete choice framework. Consider the individual decision of whether to apply to the job or not: $Pr(apply = 1) = Pr(U^j(\alpha_i, s_g, \theta_g, a_i, \hat{a}) + \xi_j > U^o(v_g, a_i, \overline{w}_g) + \xi_o)$,

[89] See also the discussion in Section 10

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.[90] 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 expected 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, obtained for the full sample through a linear truncated regression using 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.[91] The assumption is that the way in which these variables affect job performance is independent of being hired and treatment status. Appendix Figure A.5 compares the distribution of actual and imputed on-the-job ability for men and women. Parameter β_2 identifies the average utility weight on workplace gender composition α_i for gender g, β_3 identifies $(\theta_L - v_g)$, β_4 identifies the difference $(\theta_H - \theta_L)$ at mean ability level and β_5 identifies $\Delta\theta_g = \theta_H - \theta_L$.[92]

Figure 12 shows the distribution of 5000 bootstrap replications of the key coefficients β_2 and β_5, estimated separately for men and women using samples of the same size (N=800). The women's distribution of parameter α is shifted to the right of men's, indicating a stronger taste for working with own gender (or related attributes). On average, α equal 0.21 for women and -0.09 for men. The ratio of coefficients β_1 and β_2 indicates that the increase in own gender share generated by the treatment has the same effect on women's applications as a 30% increase in job wage w (an increase in the hourly wage from 16.5 to 21.45 GBP). The estimated average α for women masks heterogeneity depending on the information received and is reduced to 0.08 when estimated conditionally on low expected returns to ability. The right graph of Figure 12 shows that men's distribution of $\Delta\theta$ is shifted to the right of women's, indicating a stronger updating of expected returns to ability by the job minority. The mean $\Delta\theta$ for women is 0.01 and for men is 0.032, which implies that the ratio of priors' uncertainty between men and women is greater than one. The estimated difference of 0.032 in expected returns to ability for men is substantial: just above mean ability, it is comparable to a 16.6% increase in the wage in the job (an increase in the hourly wage from 16.5 to 19.24 GBP).

Figure A.4 shows predicted margins. For both men and women, the probability of applying is increasing in predicted on-the-job performance in the treatment with expectations of high returns to ability, but decreasing in the alternative information treatment. The differently-signed gradients indicate that posteriors on returns to ability in the job are above and below returns to ability in the outside option, respectively, in the two treatment groups ($\theta_H - v_g > 0$ and $\theta_L - v_g < 0$).

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

[91] Data on ranking and average completion rate of 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: tables, hesa and nss.

[92] The corresponding likelihood function is $lnL = \sum_{j\in S} lnF(x_j b) + \sum_{j\notin S} ln(1 - F(x_j b))$, where S is the set of all observations j, such that application outcome $y_j \neq 0$ and $F(z) = \frac{e^z}{(1+e^z)}$.

10 Alternative mechanisms

10.1 Social comparison

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. Evidence from auxiliary online experiments (Appendix B) show indeed that respondents think that the proportion of applicants with the potential of being high-achievers in the job is lower when they received the 66% than 89% statistic. Models of tournament entry (Niederle and Vesterlund, 2007; Cotton et al., 2014), directed search (Wright et al., 2017; Belot et al., 2018) and competition for jobs (Lazear et al., 2018) help us think through this alternative channel. For instance, Belot et al. (2018) show that job posts featuring higher wages, ceteris paribus, might receive a lower number of applications. This is driven by low-skilled people expecting that the earnings level will attract a more skilled pool of applicants and, thus, more competition. In the affirmative action model by Cotton et al. (2014), low ability students are discouraged by an increase in the expected degree of competition in a colour-blind contest. Similarly, in Lazear et al. (2018) low ability applicants are negatively affected by "bad luck" when competing against more skilled candidates for the same position. These models predict that we should expect low ability people not to apply when receiving information of an outstanding past performance. This would imply, consequently, an increase in average quality in the treatment featuring the 89% statistic. This contrasts with the evidence I have shown here.

10.2 Attention

Photographs may differ in the extent to which they capture the agent's attention (Gabaix, 2017).[93] In turn, only attentive agents update priors according to the information in the ad. I build an empirical measure of attention to explore this channel (Mas and Pallais, 2017). The experimental intervention was located right below a unique candidate number, which is needed to access the application portal. There are two options if a person forgets this number: searching back in their inbox for the invitation-to-apply email or asking for a reminder. Requests for reminders can be used as a proxy for endogenous attention to the intervention because candidates who asked for fewer reminders have either paid more attention to the original invitation-to-apply email or have looked back for it several times.

According to this measure, men pay relatively more attention to the female than the male photograph (see Table A.8). This is in contrast with both models of salience (Bordalo et al., 2013, 2017) and studies that would predict higher attention through perceived similarity (Forehand and Deshpandé, 2001; Jaffe, 1991). However, this evidence is similar to the results found by Bertrand et al. (2010), who show that women's photographs trigger an affective reaction which induces greater demand for credit. In my experiment, the positive reaction to the female photograph is concentrated among men in the non-white group, who received the photograph with the highest rating in terms of attractiveness (see Section B). This suggests visceral influences might also play a role in explaining job applications (see also Loewenstein, 1996). For women, demand for reminders is not different between the two photographs, which rules out that attention could be driving their behaviour.

[93]I focus on the case in which attention is not optimally chosen.

10.3 On-the-job dating market

Suppose that the main driver of choice is finding a partner on the job. Given that this is a female-dominated occupation, we should expect the proportion of single men who apply to be higher than the proportion of single women. The proportion of male applicants who are married or in a civil partnership is 19% while the proportion of married female applicants is only 12%.[94] On average, men seem not to be motivated by the on-the-job marriage market reasons. Nevertheless, the particular types of single men and women that apply for this job could still be motivated by the possibility of finding a dating opportunity. If this is the case, they might interpret employer's recruitment messages in terms of dating opportunities. This hypothesis has some testable implications.

First, we should expect heterosexual and non-heterosexual people to react in opposite ways to the same photograph. Secondly, among heterosexual people, the positive effects of seeing a person of the opposite gender should be weaker for married people.[95] Table A.9 tests these predictions. The first two columns show results for women and the last two columns for men. Columns (1) and (3) test for differential treatment effects by sexual orientation, Columns (2) and (4) by marital status. Overall, the data do not seem to support the on-the-job dating channel. A picture of the same (different) gender increases applications the most for heterosexual (non-heterosexual) women. Both these results are in contrast with the tested hypothesis. For men, there are no significant differences based on marital status. However, Column (3) shows that heterosexual men react positively to the woman's picture and negatively to the man's picture, and vice versa for non-heterosexual men. These facts are aligned with the tested hypothesis, but the effect is too small to be able not to reject this hypothesis. Moreover, the fact that non-heterosexual men and women both react positively to the non-stereotypical photograph suggests that this is not about dating.[96]

10.4 Gender differences in preferences

Competing explanations of my results based on gender differences in preferences need to account not for a simple difference in updating by men and women, but the fact that women's updating of returns to ability is affected by gender composition. I consider risk aversion, overconfidence and competitiveness (for reviews see Croson and Gneezy, 2009; Bertrand, 2011).

A wealth of studies show that men tend to be less risk averse than women (Holt and Laury, 2002; Dohmen et al., 2005; Eckel and Grossman, 2008). However, risk aversion can account for the observed behaviour only if we assume that women's risk preferences or their perception of riskiness are a function of contextual factors (for instance, the male photograph might trigger negative emotions which affect risk evaluation). Even if some studies show that women anticipate negative outcomes with greater fear than men (e.g., Fujita et al., 1991; Brody, 1993; Lerner et al., 2003), I don't have data on women's emotions to test this hypothesis. More importantly, more variance in past success

[94] The gender difference in marriage rates survives when I control for age.

[95] This is based on the assumption that faithfulness is a value for a non-zero proportion of married people.

[96] We might also think of a more complicated model of decision making, in which the information associated with a certain photograph signals the quality of potential partners. For instance, a non-heterosexual man could interpret the ad portraying a man with low performance as saying that the quality of potential partners' in the job is low. This is not what happens in the experiment, as the highest application rate for non-heterosexual men is in the treatment combining the male photo and low past performance.

does not necessarily imply higher uncertainty if people know their own ability.

Another stream of work shows that both men tend to be more overconfident than women (Lundeberg et al., 1994; Beyer and Bowden, 1997; Beyer 1998; Niederle and Vesterlund, 2007; Grosse and Reiner, 2010; Dreber et al., 2014; Coffman, 2014; Bordalo et al., 2019). However, the realm in which overconfidence is assessed matters (Jakobsson et al., 2013; Coffman et al., 2019) and the gender gap shrinks or even reverses in typical female domains. I check this in a subsample of my experimental participants (N=633). I ask them to rate themselves in ten skills on a scale from 1 (max) to 10 (max). The skills are both general (e.g., critical thinking, creativity, adaptability) and job specific (e.g., empathy, client support). For each person, I construct a measure of overconfidence by counting the number of skills rated above the sample mean.[97] Table A.10 shows that men in my sample tend to be less overconfident than women, especially in job-specific skills. Appendix I reports results from an additional exercise where I show 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 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 (Gneezy and Rustichini, 2004; Niederle and Vesterlund, 2007; Datta Gupta et al., 2013) 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., 2014).[98] This interpretation presupposes that beliefs about the returns to ability must have changed, otherwise people would have no reason to get competitive. Thus my main interpretation of the information treatment is still needed for preferences for competition to contribute to explain the results.

I then 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.[99] Figure A.6 shows that both men and women react similarly to the expectations treatment independently of this proxy of competitiveness, suggesting that competitive attitudes might not be a main driving force of the results. I further explore whether women who attended a single sex school react differently to the information manipulation, but find no evidence for this.[100]

[97] Invitation to the survey was sent to everyone in the invitation-to-apply email and subsequently encouraged through an ad-hoc email adding monetary incentives. This subsample is made of all the people that participated in the survey. The survey sample is representative of the overall pool of candidates in the field (e.g., balanced on gender, treatment assignment, FTE status).

[98] The intervention didn't change the structure of incentives on the job, as in Flory et al. (2014). People know that their earnings will not depend on their ranking. In qualitative interviews with participants, I asked them to indicate the extent to which some words came to mind by looking at the intervention email. The word "competition" had an average answer of 4.5 out of ten for both performance statistics. Moreover, if interest in a female-dominated job is negatively correlated with competitiveness, we should expect people who self-select into my experimental sample to be less competitive than average.

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

[100] Single-sex education has been shown to mitigate the gender gap in competitive attitudes by some studies, but results are ultimately mixed (Booth and Nolen, 2012; Lee et al., 2014; Laury et al., 2019). I find that women from single-sex schools (14%) react negatively to information of high expected returns to ability and even more so when combined with a female photograph (effect of 12 percentage points). If sex schooling makes women more tolerant of competition, these results go against an interpretation of the treatments in terms of competitiveness on the job. Information on schools come from the Department for Education and is available for 72% of the sample.

11 Discussion

Is attracting more men by raising expected returns to ability a free lunch? Information of higher impact of ability attracts more and better men, but it does not affect women's applications on average. In turn, a higher perceived male share improves women's selection. This suggests that breaking down barriers to men's entry into female-dominated jobs may improve the overall quality of the workforce in a gender-neutral way, ameliorating both men's and women's selection.

Table 9 provides evidence of this free lunch. The male photograph entails a decrease in applications by women as well as a slightly higher offer and acceptance rate by men. As women are better in this treatment and represent more than 75% of the workforce, their better quality drives an improvement in the overall workforce quality. Providing information of high expected returns to ability achieves the second-best overall performance of the workforce and the highest men/women gender ratio. Table 9 suggests that, in the short-term, the entry of men in female-dominated jobs might improve their selection without substantial negative spillovers on women. Once male shares increase, an improvement in women's selection might follow.

What do my results imply for talent allocation in the aggregate economy? In a world with two sectors (e.g., social and private), this ultimately depends on the nature of men's 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. This comes from the fact that switchers are the ones with 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, as my evidence suggests, 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 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 into 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.

12 Concluding remarks

Blue-collar employment is shrinking across the developed world (Autor et al., 2013; Autor et al., 2018). 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 (Autor et al., 2018; Coile and Duggan, 2019). At the same type, 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 let 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 and they grew up under traditional gender norms. Men with expectations of higher returns to ability are more likely to be hired by the employer and perform better once on the job. At the same time, a higher male share discourages the entry of less talented women in the job.

My paper assumes that men and women only differ in terms of the information they are endowed with. This contrasts with many studies on gender differences in preferences (for a review see Bertrand, 2011) and moves the focus of research from natural to nurtural differences which emerge as a result of being the minority in a certain occupation. This implies that the impact of raising expectations of returns to ability should similarly affect women's entry in male-dominated occupations. Preliminary results from a pilot experiment I conducted on an online platform seem to confirm this. I sent 900 invitations for a web development job to freelancers listed on the website, of whom women represent 30 percent. I randomized the content of the invitation in the same manner as in the field experiment presented in this paper. I replicate the result that higher expected returns to ability in the job attract more applications (7.5% compared to an average of 5%) and find that this is especially the case for women (but not statistically significant, see Table 10).

My results offer an optimistic view on the possibility to affect career choices among adults who completed their education and, in some cases, already spent several years in the labour market.[101] This is crucial for the future of work, which involves workers' quick adaptation to new roles (OECD, 2019). A related message is that recruitment strategies have the potential to reduce occupational gender segregation. 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 in supply-short male jobs during WWII (Honey 1984; Milkman 1987). This legacy inspired recent attempts to attract men in 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.

Many 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? Are men's expectations of social stigma correct? Hopefully answers to these questions may prevent communities as the ones in the Rust Belt or the North of England from being left behind by a rapidly evolving economy.

[101] This is despite the evidence that men's and women's career preferences start diverging as early as in pre-school age (Liben et al., 2001; Cvencek et al., 2011; Betrand and Duflo, 2017).

13 References

Abadie, A. (2018). *Statistical non-significance in empirical economics.* NBER Working Paper w24403. National Bureau of Economic Research.

Abebe, G., Caria, S.A., & Ortiz-Ospina, E. (2019). *The Selection of Talent Experimental and Structural Evidence from Ethiopia.* Working paper.

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics* (vol. 4, pp. 1043-1171). Elsevier.

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

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of economic perspectives*, 19(1), 9-32.

Akerlof, R. (2017). Value formation: The role of esteem. *Games and Economic Behavior*, 102, 1-19.

Altonji, J. G., & Blank, R. M. (1999). Race and gender in the labor market. *Handbook of labor economics*, 3, 3143-3259.

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., & Lee, S. (2019). Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review*, forthcoming.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, 103(6), 2121–2168.

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

Azmat, G., & Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments?. *Labour Economics*, 30, 32-40.

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

Bell, A., Chetty, R., Jaravel, X., Petkova, N., & Van Reenen, J. (2019). Who becomes an inventor in America? The importance of exposure to innovation. *The Quarterly Journal of Economics*, 134(2), 647-713.

Belot, M., Kircher, P., & Muller, P. (2018). *How Wage Announcements Affect Job Search Behavior—A Field Experiment.* IZA Working paper 11814. IZA.

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

Benyishay, A., Jones, M. R., Kondylis, F., & Mobarak, A. (2016). *Are gender differences in performance innate or socially mediated?.* Working Paper 7689. The World Bank.

Bertrand, M. (2011). New perspectives on gender. In *Handbook of labor economics* (Vol. 4, pp. 1543-1590). Elsevier.

Bertrand, M., & Duflo, E. (2017). Field Experiments on Discrimination. In *Handbook of Economic Field Experiments*, (vol. 1, pp. 309-393). North-Holland.

Bertrand, M., & 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.

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

Bertrand, M., Kamenica, E., & Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2), 571-614.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & 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.

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

Beyer (1998). Gender Differences in Self-Perception and Negative Recall Biases. *Sex Roles*, 38(1), 103–33.

Beyer, S., & Bowden, E. M. (1997). Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias. *Personality and Social Psychology Bulletin*, 23(2), 157–72.

Binder, A. J., & Bound, J. (2019). The Declining Labor Market Prospects of Less-Educated Men. *Journal of Economic Perspectives*, 33(2), 163–190.

Blau, F. D., & Kahn, L. M. (2000). Gender differences in pay. *Journal of economic perspectives*, 14(4), 75-99.

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

Blau, F. D., Brummund, P., & Liu, A. Y. H. (2013). Trends in occupational segregation by gender 1970–2009: Adjusting for the impact of changes in the occupational coding system. *Demography*, 50(2), 471-492.

Blau, F. D., Ferber, M. A., & Winkler, A. E. (2013). *The economics of women, men and work.* Pearson Higher Ed.

Blau, F. D., Simpson, P., & Anderson, D. (1998). Continuing progress? Trends in occupational segregation in the United States over the 1970s and 1980s. *Feminist economics*, 4(3), 29-71.

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

Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2), 542-555.

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

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

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

Bordalo, P., Gennaioli, N., & Shleifer, A. (2017). *Memory, attention, and choice* NBER Working Paper w23256. National Bureau of Economic Research.

Brewer, M. B., & Hewstone, M. E. (2004). *Self and social identity.* Blackwell publishing.

Brody, L. R. (1993). On Understanding Gender Differences in the Expression of Emotion. In S. L. Ablon, D. Brown, E. J. Khantzian, & J. E. Mack (eds.) *Human Feelings: Explorations in Affect Development and Meaning*, (pp. 87-121). Hillsdale, N.J.: Analytic Press.

Bureau of Labor Statistics (2019). Occupational Employment Statistics program, Occupational Projections Data. Retrieved from: https://data.bls.gov/projections/occupationProj

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

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2018). *Misperceived social norms: Female labor force participation in Saudi Arabia.* NBER Working Paper w24736. National Bureau of Economic Research.

Buser, T., Peter, N., & Wolter, S. (2017). *Gender, Willingness to Compete and Career Choices along the Whole Ability Distribution.* IZA Working paper 10976. IZA.

Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate governance, board diversity, and firm value. *The Financial Review*, 38, 33–53.

Charles, K. K., Guryan, J., & Pan, J. (2018). *The Effects of Sexism on American Women: The Role of Norms vs. Discrimination.* NBER Working Paper w24904. National Bureau of Economic Research.

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

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM?. *Social Psychological and Personality Science*, 2(6), 656-664.

Choi, J. J., Haisley, E., Kurkoski, J., & Massey, C. (2012). *Small cues change savings choices.* NBER Working Paper w17843. National Bureau of Economic Research.

Coffman, C., Collis, M., & Kulkarni, L. (2019). *Stereotypes and Belief Updating.* HBS Working Paper 19-068, Harvard Business School.

Coffman, K. B. (2014). Evidence on Self-Stereotyping and the Contribution of Ideas. *The Quarterly Journal of Economics*, 129(4), 1625–60.

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

Coffman, L., Featherstone, C. R., & Kessler, J. B. (2015). A model of information nudges. URL https://pdfs.semanticscholar.org/53ad.pdf

Coile, C. C., & 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., & Pan, J. (2018). Occupation and Gender.*The Oxford Handbook of Women and the Economy*, 425.

Cotton, C., Hickman, B. R., & Price, J. P. (2014). *Affirmative action and human capital investment:*

Evidence from a randomized field experiment. NBER Working Paper w20397. National Bureau of Economic Research.

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2), 448–474.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child development*, 82(3), 766-779.

Dal Bó, E., Dal Bó, P., & 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., & 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.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of experimental social psychology*, 40(5), 642-658.

Datta Gupta, N., Poulsen, A., & Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, 51(1), 816-835.

Del Carpio, L., & Guadalupe, M. (2018). *More Women in Tech? Evidence from a field experiment addressing social identity.* Working Paper No. 13234. CEPR Discussion Papers.

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., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2005). *Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey.* Working Paper 1730, Institute for the Study of Labor.

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

Drexler, A., Fischer, G., & Schoar, A. (2014). Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2), 1-31.

Dustmann, C., & Preston, I. (2001). Attitudes to ethnic minorities, ethnic context and location decisions. *The Economic Journal*, 111(470), 353-373.

Eckel, C. C., & Grossman, P. J. (2008). Men, Women and Risk Aversion: Experimental Evidence. In C. Plott and V. Smith *Handbook of Experimental Economics Results* (vol. 1, pp. 1061-1073), New York Elsevier.

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

Eden, D., & Ravid, G. (1982). Pygmalion versus self-expectancy: Effects of instructor and self-expectancy on trainee performance. *Organizational Behavior and Human Performance*, 30(3), 351-364.

Eisenberger, N. I., Lieberman, M. D. & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science*, 302(5643), 290-292.

Erhardt, N. L., Werbel, J. D. & Shrader, C. B. (2003). Board of director diversity and firm financial performance. *Corporate Governance*, 11(2), 102–111.

Fiske, S. T., & Taylor, S. E. (1991). *Social Cognition.* New York: McGraw-Hill.

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

Flabbi, L., Macis, M., Moro, A., & Schivardi, F. (2019). Do female executives make a difference? The impact of female leadership on gender gaps and firm performance. *The Economic Journal*, 129(622), 2390-2423.

Flory, J. A., Leibbrandt, A., & List, J. A. (2014). 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.

Forehand, M. R., & 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.

Fryer Jr, R. G., & Torelli, P. (2010). An empirical analysis of 'acting white'. *Journal of Public Economics*, 94(5-6), 380-396.

Fujita, F., Diener, E., & Sandvik, E. (1991). Gender Differences in Negative Affect and Well-Being: The Case for Emotional Intensity. *Journal of Personality and Social Psychology*, 61(3), 427–434.

Gabaix, X. (2017). *Behavioral inattention.* NBER Working Paper w24096. National Bureau of Economic Research.

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation.* WW Norton.

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

Gneezy, U., & Rustichini A. (2004). Gender and Competition at a Young Age. *American Economic Review*, 94(2), 377–381.

Gneezy, U., Niederle, M. & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3), 1049-1074.

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., & 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., Nosek, B. A. & Banaji, M. R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216.

Grosse, N. D., Riener, G., & Dertwinkel-Kalt, M. (2014). *Explaining Gender Differences in Competitiveness: Testing a Theory on Gender-Task Stereotypes.* Working Paper 017. Jena Economic Research Papers.

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

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

Handel, B., & 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.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills.*Labour economics*, 19(4), 451-464.

Hemmings, C. (2018, March 14). Men can be nursery teachers too. *BBC News.*

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

Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), 1644–1655.

Honey, M. (1984). *Creating Rosie the Riveter: Class. Gender, and Propaganda during World War II.* University of Massachusetts Press.

Hoogendoorn, S. M., & van Praag, M. (2012). *Ethnic Diversity and Team Performance: A Field Experiment.* IZA Working Paper 6731. Institute for the Study of Labor (IZA).

Hsieh, C. T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The Allocation of Talent and U.S. Economic Growth. *Econometrica*, 87(5), 1439–1474.

Imbens, G. W., & Kolesar, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, 98(4), 701-712.

Jackson, M. O. (2009). Networks and economic behaviour. *Annual Review of Economics*, 1(1), 489-511.

Jaffe, L. J. (1991). Impact of positioning and sex-role identity on women's responses to advertising. *Journal of Advertising Research*, 31(3), 57-64.

Jakobsson, N., Levin, M., & Kotsadam, A. (2013). Gender and Overconfidence: Effects of Context, Gendered Stereotypes, and Peer Group. *Advances in Applied Sociology*, 3(2), 137-141.

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

Kamas, L., & Preston, A. (2012). The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization*, 83(1), 82-97.

Kaufmann, E., & Harris, G. (2015). "White Flight" or positive contact? Local diversity and attitudes to immigration in Britain. *Comparative Political Studies*, 48(12), 1563-1590.

Kim, T., Barasz, K., John, L. K., & Morales, A. (2018). Why Am I Seeing This Ad? The Effect of Ad Transparency on Ad Effectiveness. *Journal of Consumer Research.*

Kranton, R., Pease, M., Sanders, S., & Huettel, S. (2016). *Groupy and non-groupy behaviour: Deconstructing bias in social preferences.* Working paper. W.P. Duke University.

Laury, S. K., Lee, D. J. & Schnier, K. E. (2019). Will Girls be Girls? Risk Taking and Competition in an All-Girls' School. *Economic Inquiry*, 57, 1408-1420.

Lazear, E. P. (1998). *Diversity and Immigration.* NBER Working Paper w6535. National Bureau of Economic Research.

Lazear, E. P., Shaw, K. L., & 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. *Review of Economic Studies*, 76(3), 1071–1102.

Lee, S., Niederle, M., & Kang, N. (2014). Do single-sex schools make girls more competitive?. *Economics Letters*, 124(3), 474-477.

Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of Fear and Anger on Perceived Risks of Terrorism: A National Field Experiment. *Psychological Science*, 14(2), 144–150.

Levy, G., & Razin, R. (2017). The Coevolution of Segregation, Polarized Beliefs, and Discrimination: The Case of Private versus State Education. *American Economic Journal: Microeconomics*, 9(4), 141-170.

Liben, L. S., Bigler, R. S., & Krogh, H. R. (2001). Pink and blue collar jobs: Children's judgments of job status and job aspirations in relation to sex of worker. *Journal of experimental child psychology*, 79(4), 346-363.

Loewenstein, G. (1996). Out of control: Visceral influences on behaviour. *Organizational behaviour and human decision processes*, 65(3), 272-292.

Lordan, G., & Pischke, J. S. (2016). *Does Rosie like riveting? Male and female occupational choices.* NBER Working Paper w22495. National Bureau of Economic Research.

Lundeberg, M. A., Fox, P. W., & J. (1994). Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments. *Journal of Educational Psychology*, 86(1), 114– 21.

Marinescu, I., & Wolthoff, R. (2016). *Opening the black box of the matching function: The power of words.* NBER Working Paper w22508. National Bureau of Economic Research.

Martin, C. L. (1990). Attitudes and expectations about children with nontraditional and traditional gender roles. *Sex roles*, 22(3-4), 151-166.

Marx, D. M., & 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., & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12), 3722-3759.

McCaffrey, D. F., & Bell, R. M. (2002). Bias reduction in standard Errors for linear and generalized linear models with multi-stage samples. In *Proceedings of Statistics Canada Symposium*, (pp. 1-10).

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 415-444.

Milkman, R. (1987). *Gender at work: The dynamics of job segregation by sex during World War II* (vol. 286). University of Illinois Press.

Miller, C. C. (2017, January 4). Why Men Don't Want the Jobs Done Mostly by Women. *The New York Times.*

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

Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men & Masculinity*, 11(2), 140.

Murphy, K. M., Shleifer, A., & Vishny, R. W. (1991). The allocation of talent: Implications for growth. *The Quarterly Journal of Economics*, 106(2), 503-530.

Nedelkoska, L., & Quintini, G. (2018). *Automation, skills use and training.* Working paper 202. OECD Publishing.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3), 799-866.

Ngai, L. R., & 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., & 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., & Yestrumskas, A. H. (2008). *Gender differences in seeking challenges: the role of institutions.* Working Paper.

OECD (2019). *OECD Employment Outlook 2019: The Future of Work*, OECD Publishing, Paris.

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

Oxoby, R. J. (2014). Social inference and occupational choice: Type-based beliefs in a Bayesian model of class formation. *Journal of Behavioral and Experimental Economics*, 51, 30-37.

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

Phelan, J. E., & Rudman, L. A. (2010). Reactions to ethnic deviance: The role of backlash in racial stereotype maintenance. *Journal of Personality and Social Psychology*, 99(2), 265.

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

Propper, C., & Van Reenen, J. (2010). Can pay regulation kill? Panel data evidence on the effect of labor markets on hospital performance. *Journal of Political Economy*, 118(2), 222-273.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, 111(12), 4403-4408.

Reuben, E., Wiswall, M. & Zafar, B. (2015). Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. *The Economic Journal*, 127, 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. Institute for the Study of Labor (IZA).

Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current directions in psychological science*, 3(6), 176-179.

Roy,A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers* (New Series), 3(2), 135-146.

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behaviour: the role of backlash in cultural stereotype maintenance.*Journal of personality and social psychology*, 87(2), 157.

Sandnabba, N. K., & Ahlberg, C. (1999). Parents' attitudes and expectations about children's cross-gender behaviour. *Sex Roles*, 40(3-4), 249-263.

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association*, 12(6), 1423-1452.

Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. *Advances in experimental social psychology*, 24, 161-199.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & 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., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology*, 69(5), 797.

Stevenson, M. (2018). *Demand for Healthcare Workers Will Outpace Supply by 2025: An Analysis of the US Healthcare Labor Market*. Mercer Report HPA.

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

Stone, J., Perry, W., & 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., & Turner, J. (1986). The social identity theory of intergroup behaviour. In *Psychology of intergroup relations* (vol. 5, pp. 7-24). Chicago: Nelson Hall.

The Economist (2018, August 18). A shortage of nurses calls for the recruiting of more men. But that is a tricky task. *The Economist*.

Whatley, W., & Wright, G. (1994). Race, Human Capital, and Labour Markets in American History. In G. Grantham, and M. MacKinnon, *Labour Market Evolution*, (pp. 282-303). Routledge: London and New York.

Whittingham, A. (2018, February 8). Government must act to halt crisis in social care recruitment. *The Guardian*.

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

Wiswall, M. & Zafar, B. (2018). Preference for the Workplace, Human Capital, and Gender. *The Quarterly Journal of Economics*, 133(1), 457–507.

Wozniak, D., Harbaugh, W., & Mayr, U. (2010). *Choices about competition: differences by gender and hormonal fluctuations, and the role of relative performance feedback*. MPRA Paper 21097.

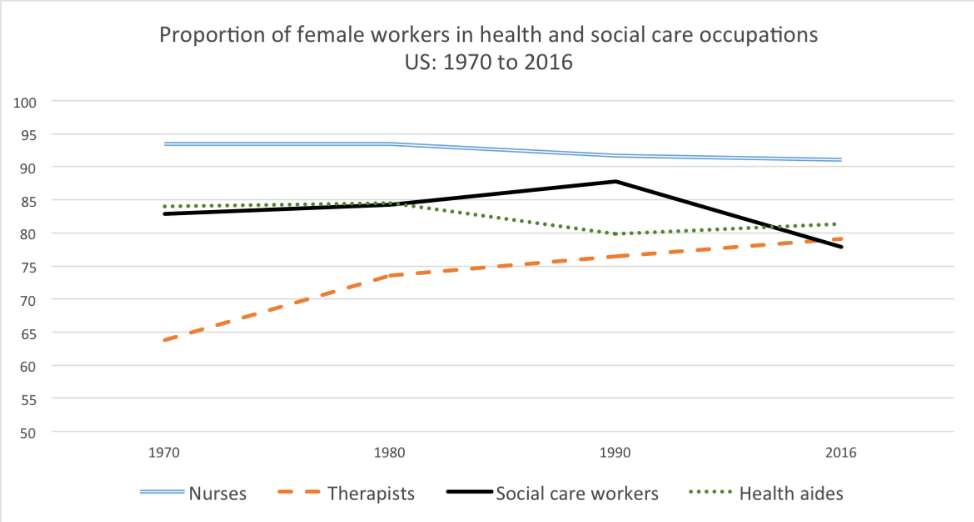
Wright, R., Kircher, P., Julîen, B., & Guerrieri, V. (2017). *Directed search: A guided tour*. NBER Working Paper w23884. National Bureau of Economic Research.

Young, A. (2018). Channelling 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, 545–595.

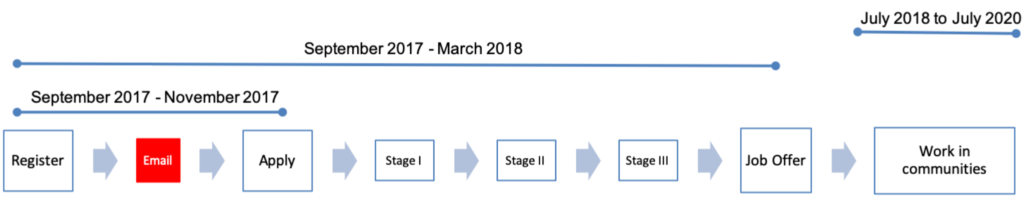
14 Figures

Figure 1. Female shares in selected occupations in the U.S.: 1970 to 2016



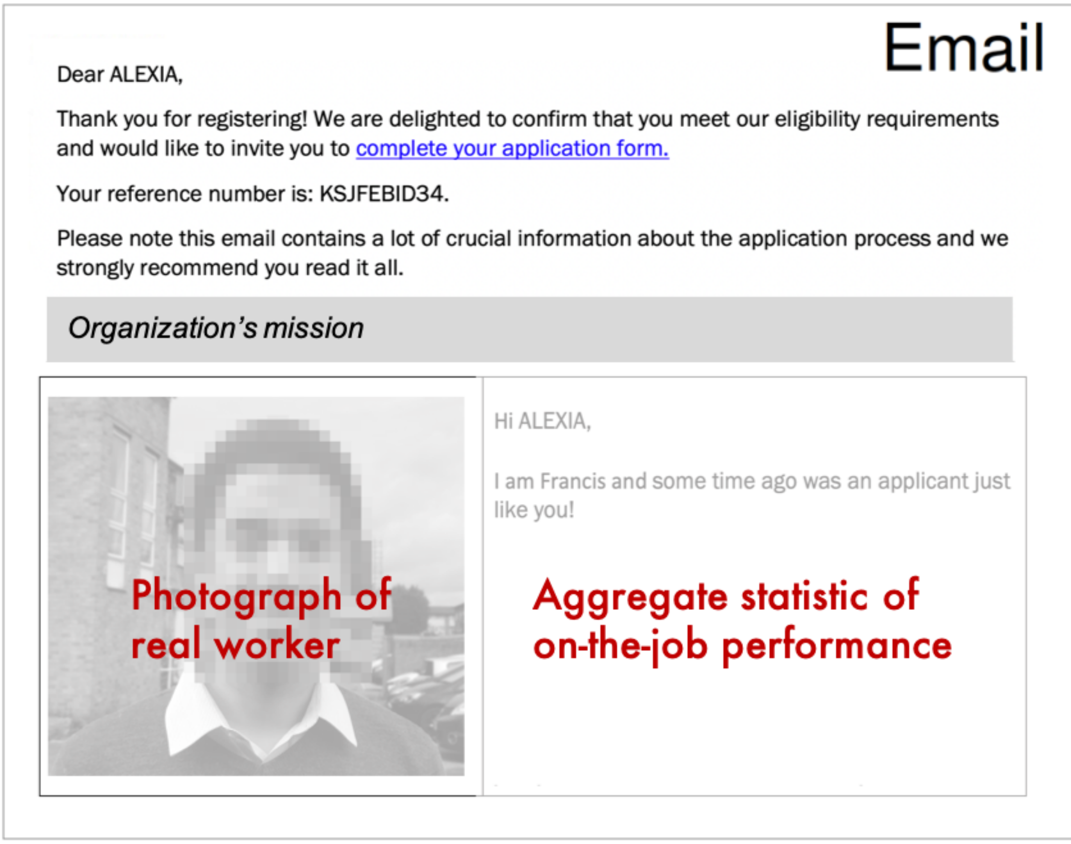
Note. Data for 1970 to 1990 are from Blau et al. (1998), who use the U.S. Census Data (U.S. Department of Commerce, Bureau of the Census) Data for 2016 are from the American Community Survey. The category "Nurses" include: licensed and registered nurses, licensed practical and vocational nurses and nursing aides. The category "Therapists" includes: occupational, physical, speech and others. The category "Social care workers" includes: social workers, childcare workers, social welfare workers, social and community service occupations/managers, community health workers. The category "Health aides" excludes nurses.

Figure 2. Recruitment timeline



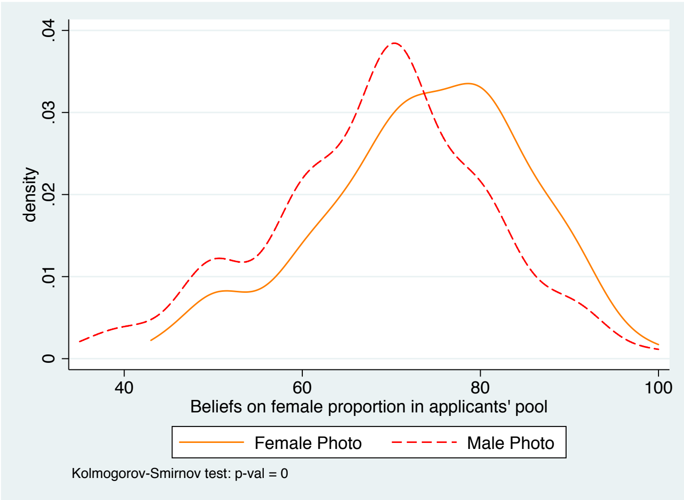
Note. The Figure shows the recruitment timeline of the partner organization from the candidates' perspective. Applications were open from September until November 2017. Randomization of the invitation-to-apply was happening between online registration and application submission. After submitting the application, the hiring process consisted of different assessment stages (e.g., interviews). If a person was hired and accepted the job, actual work in local authorities started in July 2018.

Figure 3. Intervention email template



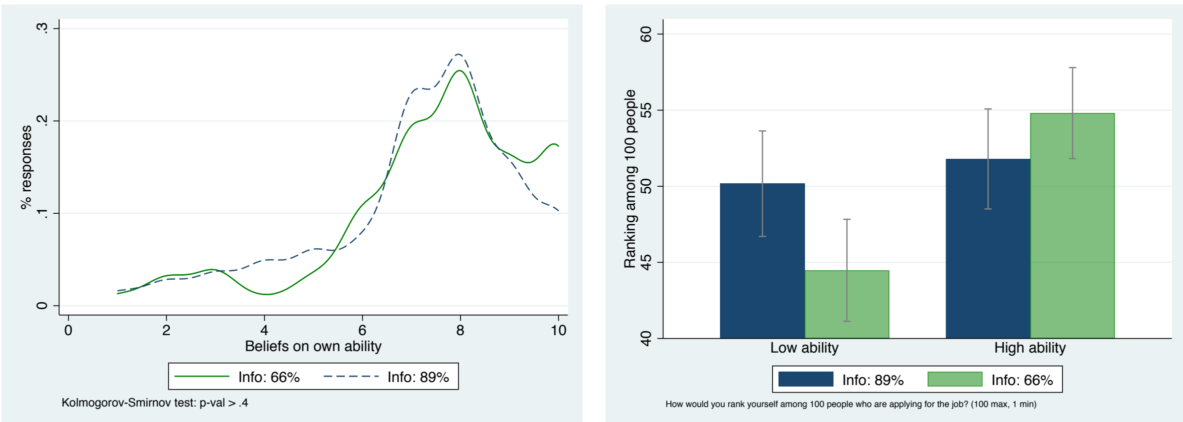
Note. The Figure shows a stylized example of one of the email templates used in the intervention.

Figure 4. Gender shares shock: manipulation checks



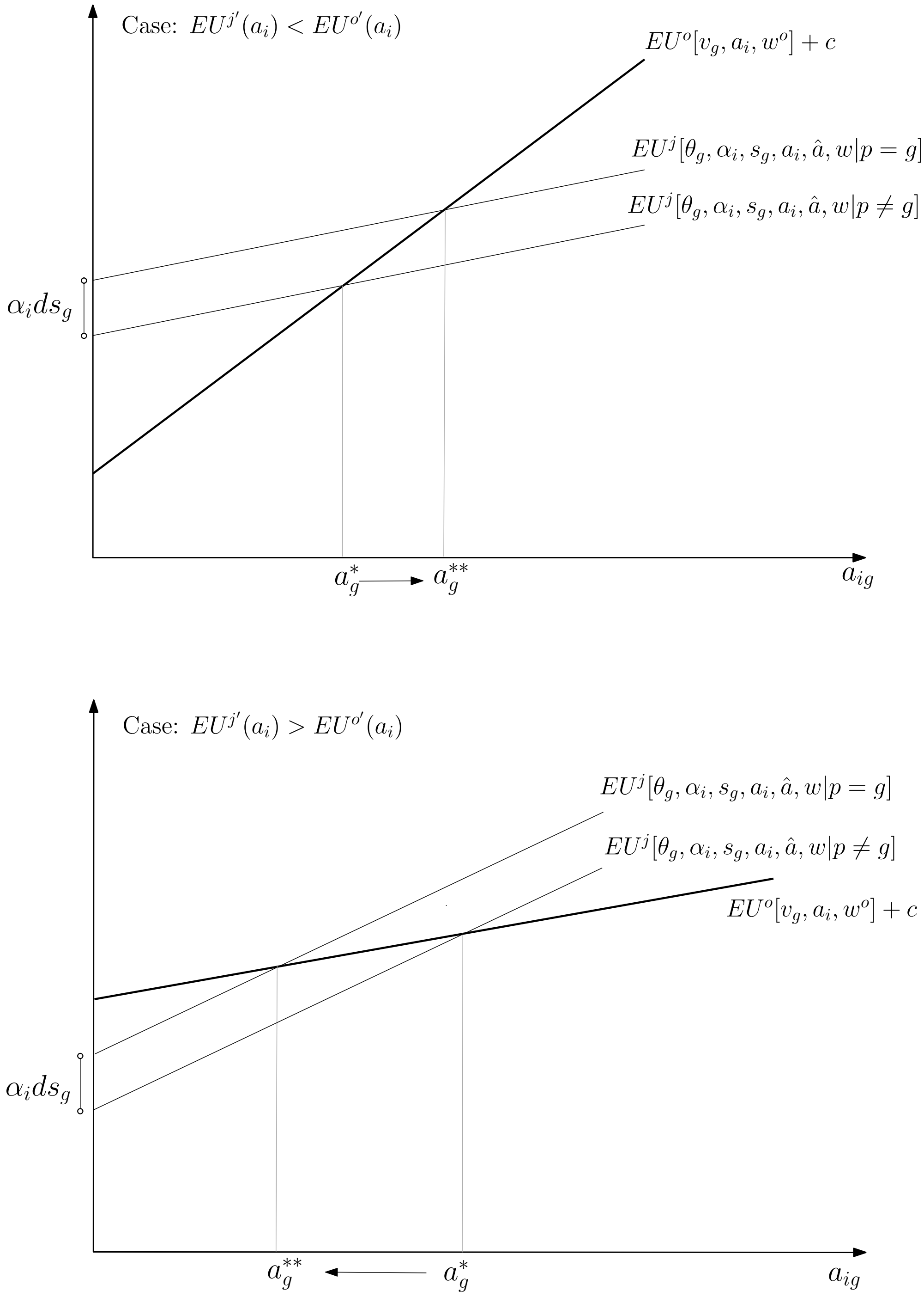
Note. The figure shows the distribution of answers to the question "Consider 100 people who apply for this job. How many do you think are women?", separately for respondents assigned to the email with a female or male photograph. Data are from the auxiliary online surveys. The dashed (solid) line is 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 5. Expectations shock: manipulation checks



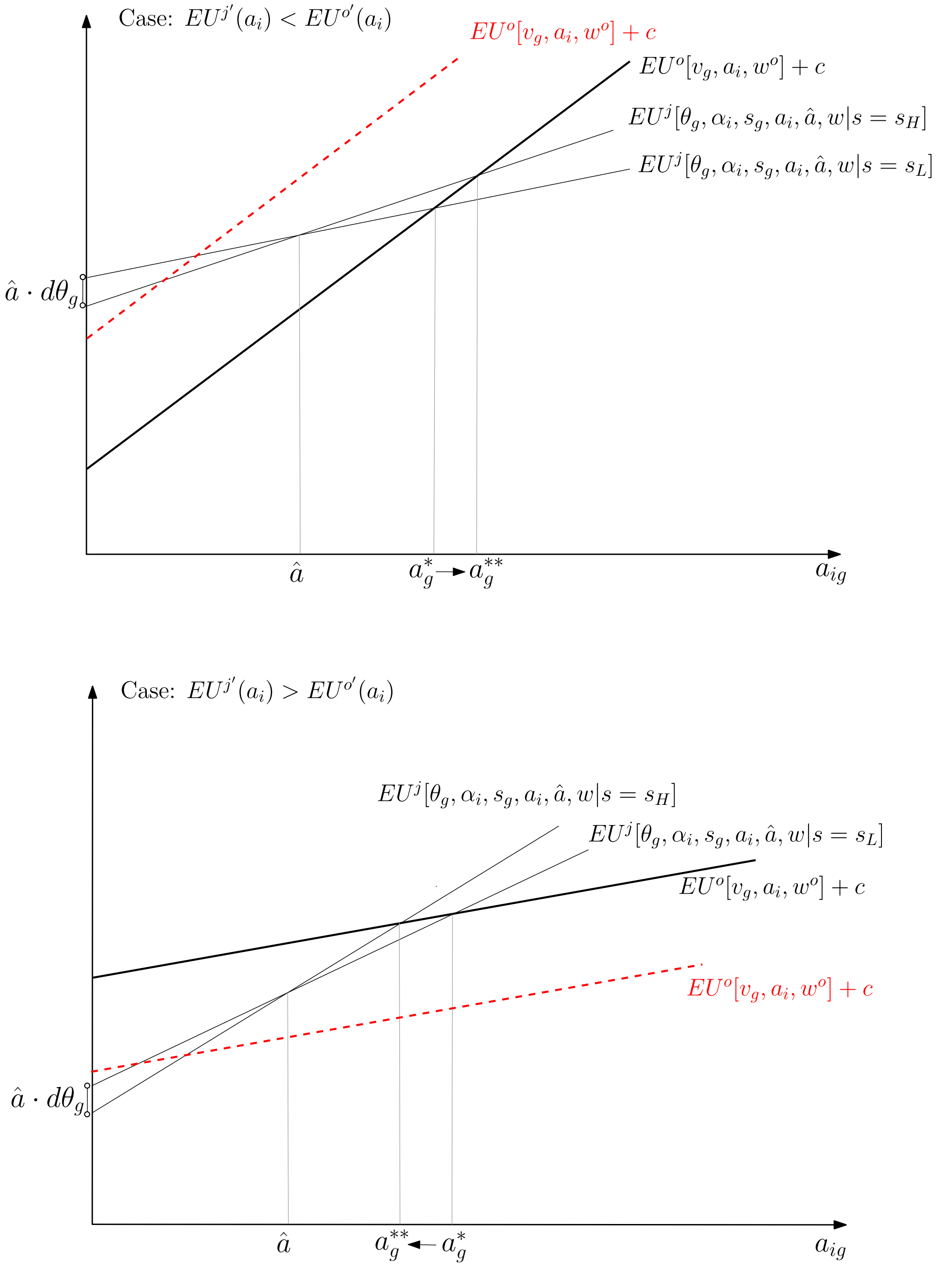
Note. The left panel shows the distribution of answers to the question "How do you expect a person with your skills and experience to perform in interacting with families in need?" on a scale from 1 (min) to 10 (max), separately for respondents assigned to the email with a statistic of 66% (solid line) 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 6. Theory: effect of a shock to perceived gender shares



Note. The figure plots the application decision for potential applicants of gender g. The top panel considers the case $U^{j'}(a_i) < U^{o'}(a_i)$ and the bottom panel the case $U^{j'}(a_i) > U^{o'}(a_i)$. The solid thick line shows the outside option. The two thin solid lines show the expected job utility when receiving a photo of the same ($p = g$) or different gender ($p \neq g$). The vertical distance between the two solid thin 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 $p \neq g$. In the top panel, the marginal applicant a_g^{**} is more skilled than a_g^*. The opposite result for quality holds in the bottom panel.

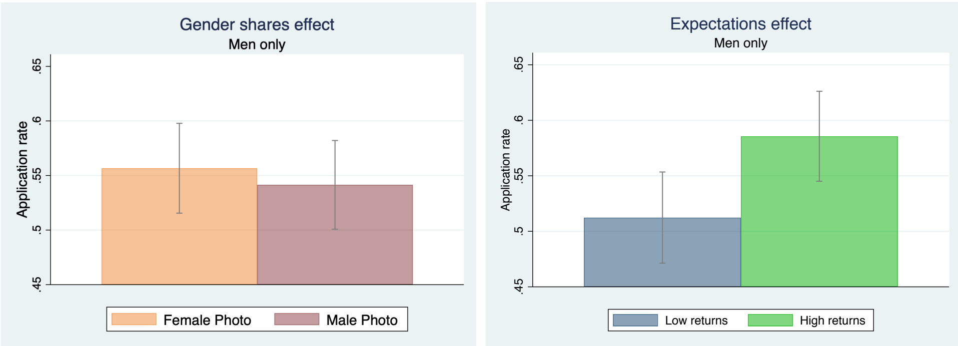
Figure 7. Theory: effect of a shock to expectations of returns to ability



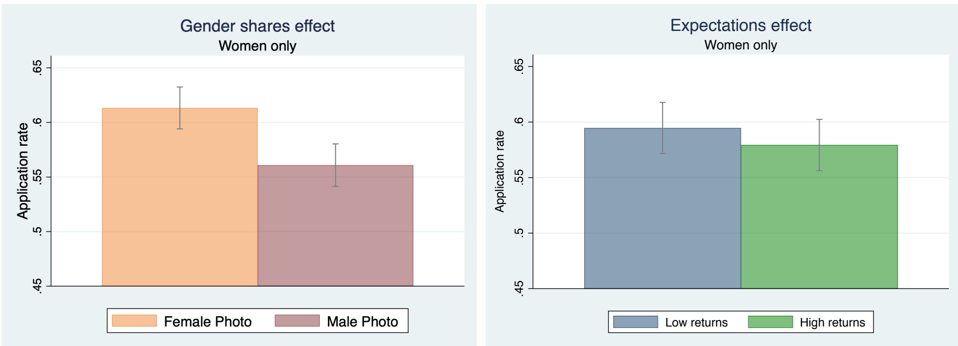
Note. The figure plots the application decision for potential applicants of gender g. The top panel considers the case $U^{j'}(a_i) < U^{o'}(a_i)$ and the bottom panel the case $U^{j'}(a_i) > U^{o'}(a_i)$. The solid black and dashed red thick lines show two different levels of the outside option. The two thin solid lines show the expected job utility when receiving information of high ($s = s_H$) or low ($s = s_L$) returns to ability. The different slope of the two solid thin lines is explained by $E[\theta|s = s_H] > E[\theta|s = s_L]$. 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 $U^{j'}(a_i) < U^{o'}(a_i)$. In the top panel, the marginal applicant a_g^{**} is more skilled than a_g^* if $B > 0$. The opposite result for quality holds in the bottom panel. The red dashed lines illustrate cases in which condition $B > 0$ is violated (i.e. $a_g^* < \hat{a}$)

Figure 8. Application rates by treatment and gender

(a) Men



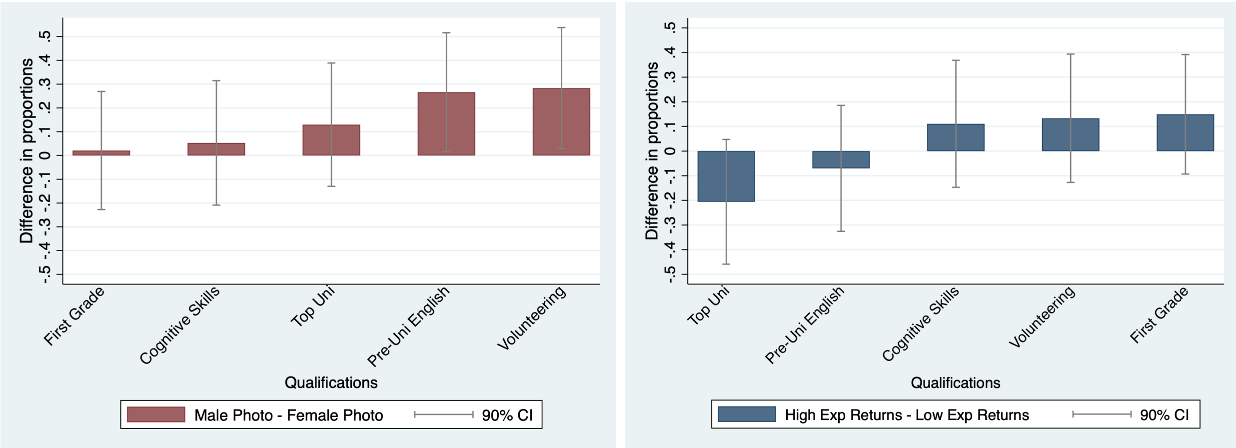
(b) Women



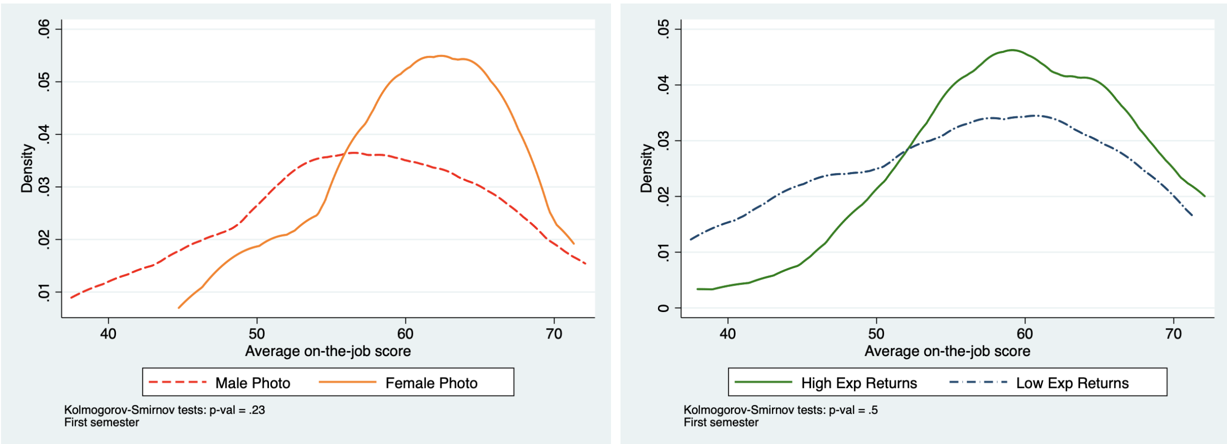
Note. Panel A shows application rates for men by photograph treatment (left-hand side) and information treatment (right-hand side). Panel B shows application rates for women by photograph treatment (left-hand side) and information treatment (right-hand side). Error bars show 95% confidence intervals.

Figure 9. Men's qualifications and average on-the-job test scores by treatment

(a) Qualifications



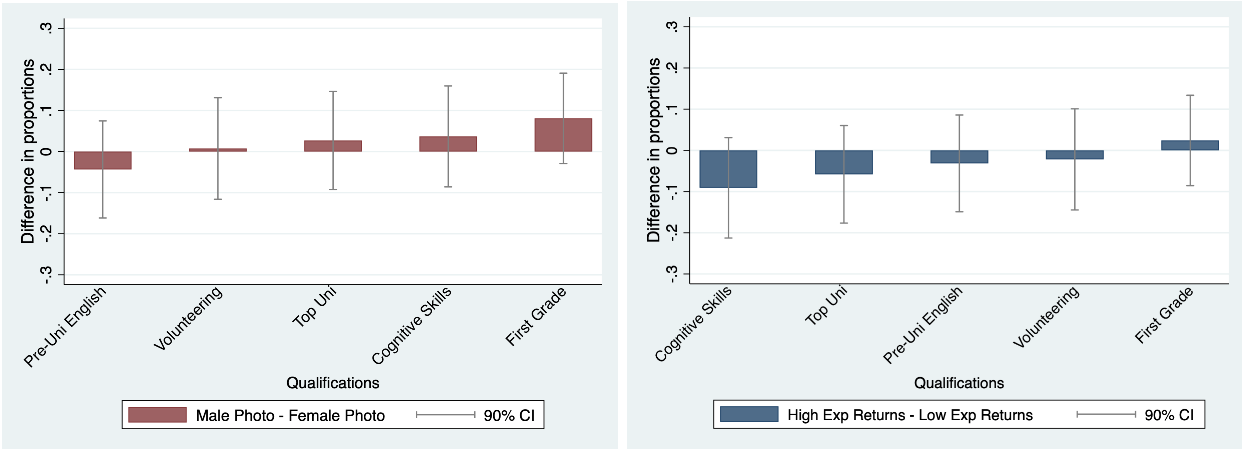
(b) Average on-the-job test scores



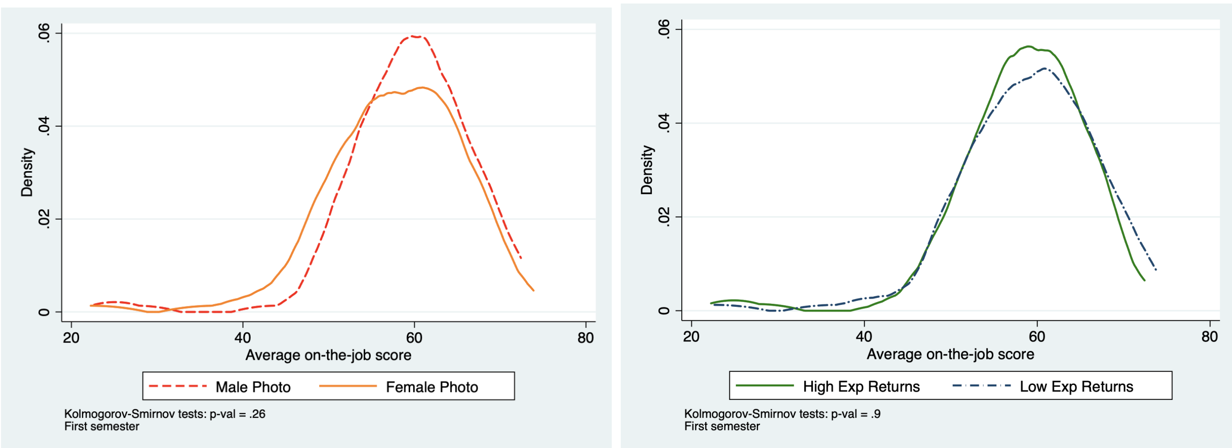
Note. The figure shows differences in the proportion of men that hold a certain qualification between treatment groups (Panel a) and the cumulative distribution of men's average test scores during the first semester on the job (Panel b). Figures on the left-hand side show the distributions by photograh 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. Workers are assessed on five different assessments on a scale from 0 (min) to 100 (max), where 50 is the minimum score for passing the test.

Figure 10. Women's qualifications and average on-the-job test scores by treatment

(a) Qualifications

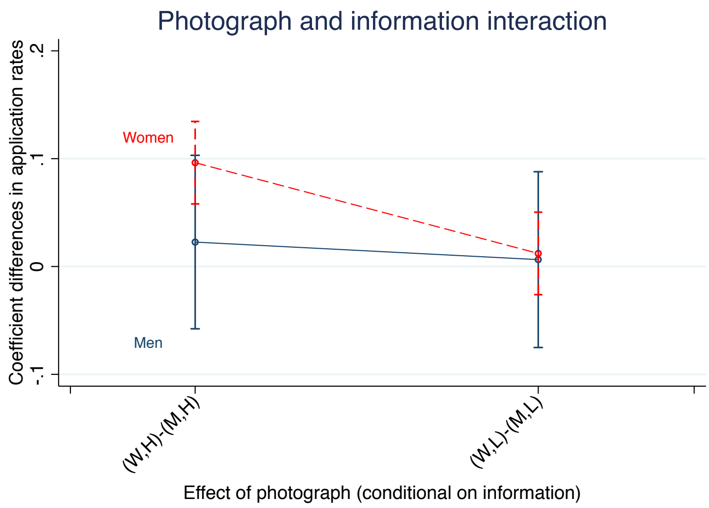


(b) Average on-the-job test scores



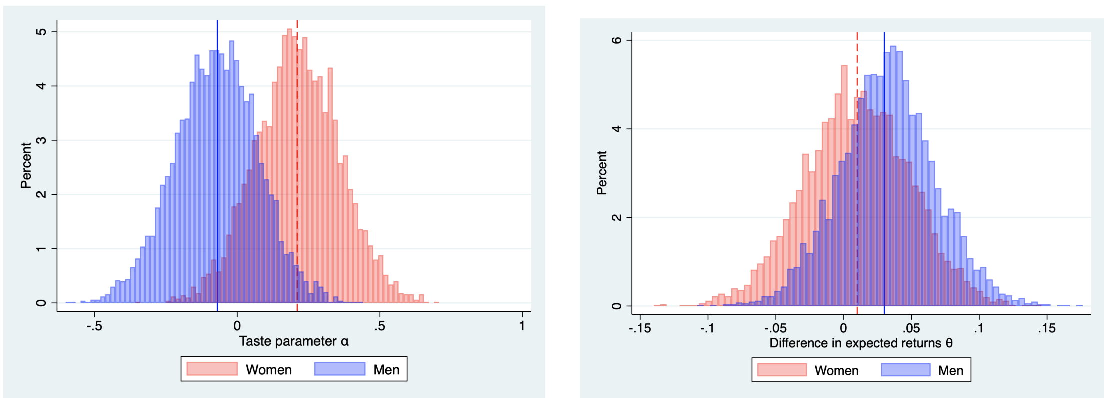
Note. The figure shows differences in the proportion of women that hold a certain qualification between treatment groups (Panel a) and the cumulative distribution of women's average test scores during the first semester on the job (Panel b). Figures on the left-hand side show the distributions by photograh 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. Workers are assessed on five different assessments on a scale from 0 (min) to 100 (max), where 50 is the minimum score for passing the test.

Figure 11. Interaction between photographs and information on applications



Note. The figure shows the effect of the treatment on application rates for each of the four experimental groups. The figure on the left-hand side shows the difference in application rates between the high and low expected returns treatments conditional on each type of photograph. That is, $(p, H) - (p, L)$ with $p \in \{M, W\}$. The figure on the right-hand side shows the 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 blue solid lines are for men.

Figure 12. Structural parameters' estimates



Note. The figure shows distributions of the estimated parameters α on the left-hand side and $\Delta\theta$ on the right-hand side. Blue bars are for men and red bars are for women. Vertical lines are the mean value of the parameters for each gender. Multiple estimations are obtained through 5000 bootstrap replications of the logit model described in the main body of the paper.

15 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 by 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 U.K. universities those belonging to the Russell Group (see here). "Graduate" is a dummy for whether the candidate graduated in 2016 or before. "Scientific Subject" assumes value one if the person studied engineering, IT/Computer Science, Maths or Natural Sciences. "Past application" is a dummy 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 submission of the application. "Early registration" is a dummy equal to one if the person had access to an early opening of the application. "Registration by November" is a dummy for whether the person started the application process before the end of October. "Any event" is a dummy equal to one if the candidate participated in any of the organization's career events. "Economic school support" is a dummy equal to 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. 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 report the minimum p-value from the associated t-test between pairs of treatment groups with robust standard errors and with pooled genders.

Table 2. Men's results

	(1)	(2)	(3)	(4)
DV:	Applied and never DO	Received Offer	Accepted Offer	Avg. Score on-the-job
Male Photo	-0.017	0.055	0.090	-0.316
	(0.035)	(0.034)	(0.124)	(0.228)
High Exp Returns	0.071**	0.061*	-0.023	0.467*
	(0.035)	(0.033)	(0.128)	(0.249)
Observations	807	440	67	43
R-squared	0.018	0.062	0.035	0.210
Basic Controls	Y	Y	Y	Y
Mean Dep Var	0.53	0.21	0.83	-0.03
Photo = Exp Ret p-val	0.08	0.89	0.53	0.02
Rand Inf p-val				
Photo	0.63	0.11	0.47	0.19
Exp Returns	0.04	0.08	0.83	0.07

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 four different regressions. The omitted category is the treatment group which received the female photograph and information of 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 of high returns to ability (specification (2) of Section 5.1). The dependent variables are indicators 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. The score is standardized by subtracting the mean and dividing by the standard deviation of the gender-specific distribution. 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. Treatment effects by exposure to gender occupational segregation

DV: Applied and never DO = 1				
	(1)	(2)	(3)	(4)
	Job Genderization		Men in Pink-Collar	
	High	Low	Low	High
Male Photo	-0.026	-0.011	-0.010	-0.020
	(0.050)	(0.050)	(0.050)	(0.050)
High Exp Returns	0.167***	-0.021	0.112**	0.036
	(0.050)	(0.050)	(0.050)	(0.050)
Observations	390	402	394	398
R-squared	0.038	0.017	0.022	0.030
Basic Controls	Y	Y	Y	Y
Mean Dep Var	0.55	0.52	0.56	0.51
Photo = Exp Ret p-val	0.01	0.88	0.08	0.43
Rand Inf p-val				
Photo	0.58	0.80	0.86	0.70
Exp Returns	0.00	0.66	0.03	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 four different regressions. The omitted category is the treatment group which received the female photograph and information of 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 of high returns to ability (specification (2) of Section 5.1). 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 live (either currently or in the past). The index is computed using data from the 2011 U.K. Census. 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 variable "Men in Pink-Collar" is the average share of men in female-dominated jobs at the local area level (MSOA) where the subject went to secondary school or live (either currently or in the past). The index is computed using data from the 2011 U.K. Census. I defined female-dominated occupations the ones that have more than 75% 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. 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.

Table 4. Treatment effects by wage dispersion and level of outside option

DV: Applied and never DO = 1					
	(1)	(2)	(3)	(4)	(5)
	Wage dispersion		Quantiles of outside option		
	Low	High	1^{st}	2^{nd}	3^{rd}
Male Photo	-0.014	-0.025	-0.060	-0.011	0.048
	(0.046)	(0.055)	(0.062)	(0.061)	(0.059)
High Exp Returns	0.036	0.122**	0.103*	0.050	0.069
	(0.046)	(0.055)	(0.061)	(0.062)	(0.059)
Observations	477	330	260	266	281
R-squared	0.012	0.033	0.029	0.029	0.061
Basic controls	Y	Y	Y	Y	Y
Mean Dep Var	0.56	0.55	0.47	0.59	0.54
Photo = Exp Ret p-val	0.44	0.06	0.06	0.49	0.79
Rand Inf p-val					
Photo	0.74	0.67	0.34	0.84	0.41
Exp Returns	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 table reports results of five different regressions. The omitted category is the treatment group which received the female photograph and information of 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 of 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 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 5. 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	Y		Y	

Robust standard errors in parentheses

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

Notes: OLS estimates. In Column (1), T^2 indicates information of 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 6. On-the-job performance: panel data

	DV: First Semester Std. Scores	
	(1)	(2)
Male Photo	-0.110	-0.255
	(0.145)	(0.193)
High Exp Returns	0.246*	0.361**
	(0.129)	(0.142)
Observations	215	215
R-squared	0.235	0.293
Basic Controls	Y	Y
Mean Dep Var	0.04	0.04

Clustered standard errors in parentheses (ind. level)

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

Note. OLS panel estimates for men only. The table reports results of two different regressions. The omitted category is the treatment group which received the female photograph and information of 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 of high returns to ability (specification (2) of Section 5.1). Column (2) introduces weights 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 score in Maths pre-university tests. Standard errors are clustered at the worker level.

Table 7. Perceived social impact and intent to stay

	(1)	(2)	(3)	(4)	(5)	(6)
DV:	Perceived impact		Confidence	Recommend	Intent to stay	
	at work	outside	own practice	programme	in LA	in job
Male Photo	0.215*	0.075	0.136	0.084	0.237	0.265**
	(0.123)	(0.154)	(0.144)	(0.111)	(0.144)	(0.107)
High Exp Returns	0.195	0.024	0.092	0.300**	0.104	0.363***
	(0.117)	(0.153)	(0.138)	(0.116)	(0.131)	(0.112)
Observations	38	38	38	38	38	38
R-squared	0.215	0.052	0.183	0.256	0.138	0.417
Basic Controls	Y	Y	Y	Y	Y	Y
Mean DV	0.60	0.20	0.67	0.87	0.60	0.87
Photo = Exp Ret p-val	0.91	0.84	0.83	0.15	0.51	0.52

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 of 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 of high returns to ability (specification (2) of Section 5.1). "Perceived impact" is an indicator equal to one if a worker feels that he is having positive social impact in his work (Column 1) or outside work (Column 2). "Confidence in own practice" is equal to one if the worker feels confident in interacting with families in need. "Recommend the programme" is equal to one if the worker would recommend the job to others. "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 (Column 5) or in the same job (Column 6). 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 8. Women's results

	(1)	(2)	(3)	(4)
DV:	Applied and never DO	Received Offer	Accepted Offer	Avg. Score on-the-job
Male Photo	-0.051***	0.013	0.131**	0.194^{+}
	(0.017)	(0.015)	(0.055)	(0.136)
High Exp Returns	-0.015	0.004	-0.002	-0.018
	(0.017)	(0.015)	(0.055)	(0.136)
Observations	3,513	2,062	301	191
R-squared	0.013	0.025	0.028	0.280
Basic Controls	Y	Y	Y	Y
Mean Dep Var	0.59	0.15	0.68	0.09
Photo = Exp Ret p-val	0.12	0.67	0.08	0.25
Rand Inf p-val				
Photo	0.00	0.40	0.03	0.15
Exp Returns	0.35	0.81	0.98	0.90

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1, $^{+}$ p<0.15

Note. OLS estimates for women only. The table reports results of four different regressions. The omitted category is the treatment group which received the female photograph and information of 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 of high returns to ability (specification (2) of Section 5.1). The dependent variables are indicators 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. The score is standardized by subtracting the mean and dividing by the standard deviation of the gender-specific distribution. 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 9. Gender ratio and on-the-job performance: summary

	M/W gender ratio			**On-the-job performance**		
	Applicants	Offerees	Workers	Women	Men	Overall
			Photograph			
Female Photo	21%	18%	19%	57.77	60.59	58.22
				[8.0]	[7.0]	
Male Photo	22%	26%	25%	59.36	56.57	58.80
				[7.6]	[9.9]	
			Information			
Low Exp Returns	20%	17%	18%	58.96	55.46	58.43
				[7.9]	[10.0]	
High Exp Returns	23%	28%	27%	58	59.77	58.66
				[7.7]	[8.2]	

Note. The first three columns of this table show the men/women gender ratio among applicants (Column 1), people who received a job offer (Column 2) and workers (Column 3). The last three columns show the average test scores achieved on the job by women (Column 4), men (Column 5) and the weighted average of these two, where weights are given by gender shares. Digits in square brackets report the standard deviation of average test scores. The average on-the-job test score is computed as the average of the first five assessments during the first semester on the job. The score is between 0 (min) and 100 (max).

Table 10. Replicating the experiment in a male-dominated sector: results from a pilot

DV: Applied = 1			
	(1)	(2) Women	(3) Men
High Exp Returns	0.023*	0.034	0.015
	(0.014)	(0.023)	(0.022)
Observations	900	408	492
R-squared	0.039	0.058	0.055
Basic Controls	Y	Y	Y
Country FE	Y	Y	Y
Mean DV	0.05	0.05	0.05

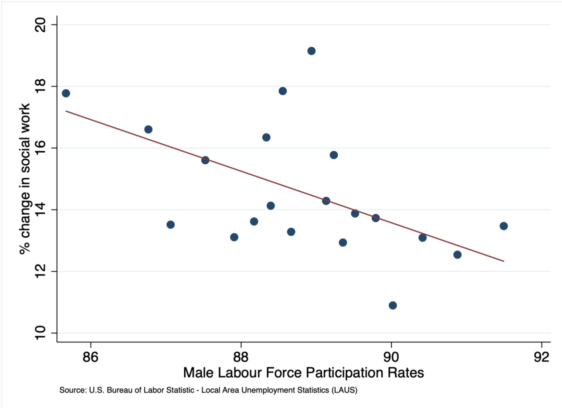
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note. Preliminary results from a pilot experiment conducted on an online platform. I sent 900 invitations for a web development job to freelancers listed on the website. The regressor “High Exp Returns” is a dummy equal to one for receiving information of high returns to ability, which in this context was “Did you know that 68% of freelancers hired for similar jobs got 4.9 or 5 stars in clients’ feedback?”. In the alternative treatment, the percentage was 87. All the regressions control for the basic set of controls X_i made of the following variables: ethnicity, gender and being above/below median posted hourly price (stratification variables), day of invitation, number of skills listed, having wed-development skills, having had less than five clients, having missing client feedback.

Appendices

A Appendix figures and tables

Figure A.1. Social work growth and male labour force participation



Note. The figure shows a binned scatterplot between the 2018 male labour force participation (on the x-axis) and employment growth 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. Expectations effect and job-specific ability

DV: Applied and never DO = 1			
	(1)	(2)	(3)
	Ability		
	Low	High	
High Exp Returns	0.041	0.101**	-0.178
	(0.049)	(0.051)	(0.653)
Ability a_i			-0.001
			(0.008)
High Exp Returns * Ability a_i			0.004
			(0.011)
Observations	410	397	807
R-squared	0.016	0.035	0.019
Basic Controls	Y	Y	Y
Mean Dep Var	0.53	0.53	0.53

Standard errors in parentheses

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

Note. OLS estimates for men only. The regressor "High Exp Returns" is a dummy equal to one for information of high expected returns to ability (specification (2) of Section 5.1). "Ability" is computed as the predicted performance on the job and takes values between 0 and 100. Predicted performance is calculated using a truncated linear regression 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. The level "high" or "low" is defined for values of the variable respectively above or below the 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.

Table A.2. 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	Y	Y
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 (specification (3) 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 indicated treatment dummies from randomization inference (randomization-t) with 1000 repetitions.

Table A.3. Do women and men react differently to treatments?

	(1)	(2)	(3)
VARIABLES	Applied and never DO	Received Offer	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	Y	Y	Y
Mean Dep Var	0.590	0.590	0.590

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 of 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 of high returns to ability (specification (2) of Section 5.1). The dependent variables are indicators 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.4. 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	Y	Y	Y	Y	Y	Y
Mean Dep Var	0.62	0.58	0.67	0.60	0.58	0.52
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 of 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 of 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.5. Treatment effects by regional wage dispersion

DV: Applied and never DO	(1)	(2)	(3)	(4)
	Men		Women	
	Wage dispersion: Low	Wage dispersion: High	Wage dispersion: Low	Wage dispersion: 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	0.55	0.57	0.620	0.590
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

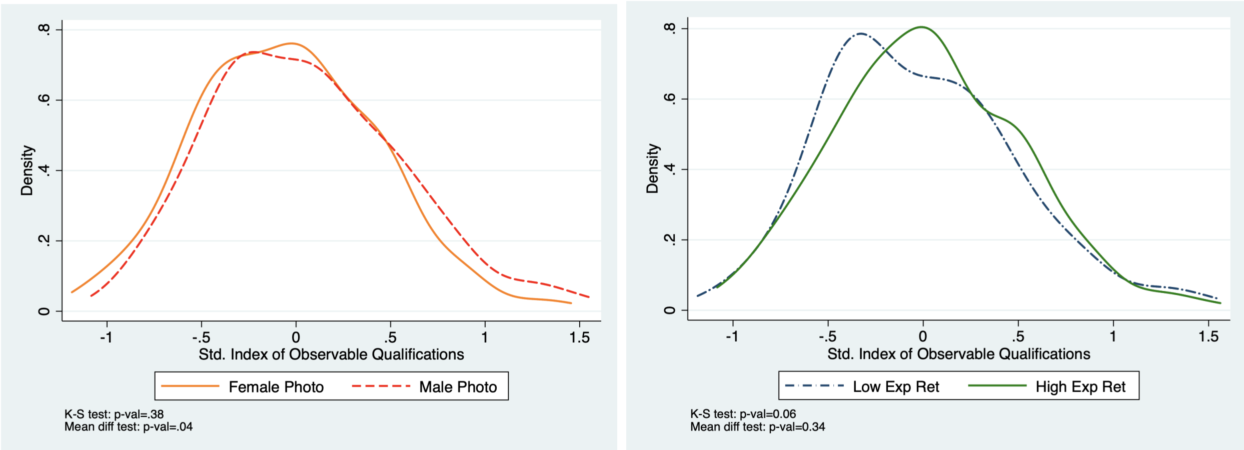
Robust standard errors in parentheses

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

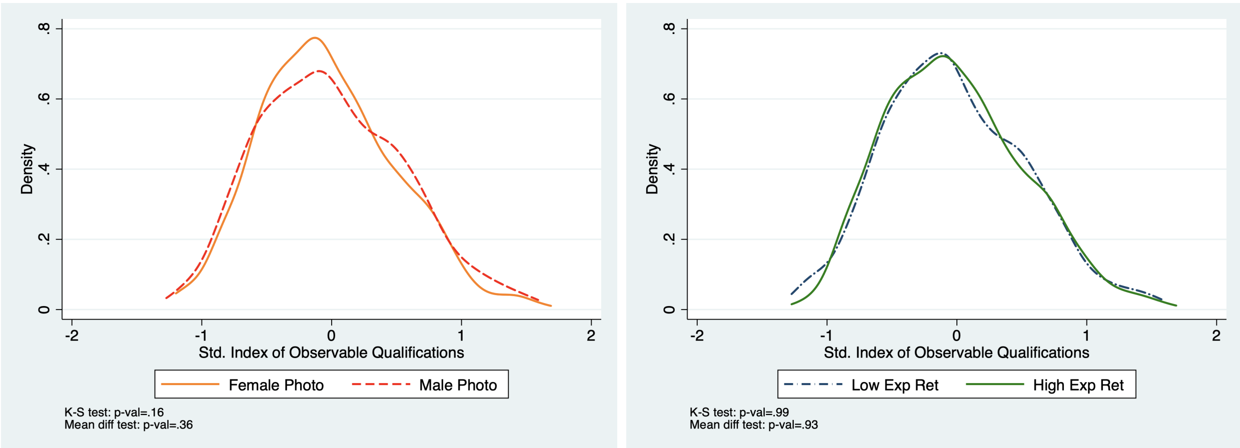
Note. OLS estimates run separately for men (Columns 1 and 2) and women (Columns 3 and 4). "Wage dispersion" is computed as the 75/25 interquartile range of the gender-specific distribution of hourly wages across industries in the UK region where the candidate lives. The level "high" or "low" is defined 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.

Figure A.2. Applicants' index of desirable qualifications by treatment

(a) Men



(b) Women



Note. The figure shows the distribution of a standardized index of desirable qualifications between treatment groups for men (Panel a) and women (Panel b). Figures on the left-hand side show the distributions by photograh 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 low expected returns to ability. The index is computed as the average 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.

Table A.6. Effort in application completion

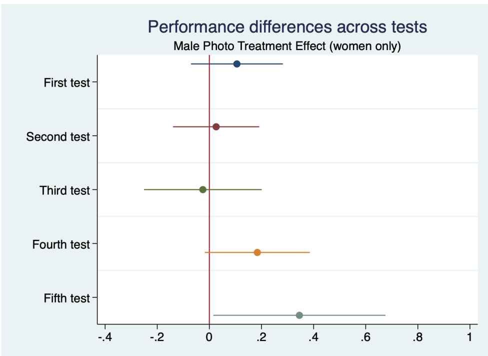
VARIABLES	(1) Access to portal	(2) # edits	(3) % completed	(4) Qst 1 length	(5) Qst 2 length
High Exp Returns	0.009	4.373**	0.026	34.178	42.509
	(0.025)	(1.970)	(0.023)	(55.655)	(46.081)
Observations	804	807	807	807	807
R-squared	0.033	0.043	0.030	0.022	0.027
Basic Controls	Y	Y	Y	Y	Y
Week dummies	Y	Y	Y	Y	Y

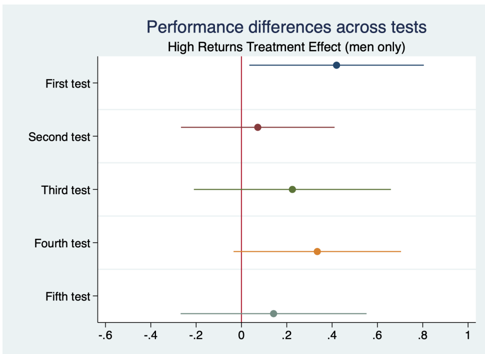
Robust standard errors in parentheses

*** 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 of 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 of 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.3. On-the-job test scores differences by treatment over time





Note. The figure reports the coefficients from a regression of each of the five 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). The left figure is for women only and the right figure for men only. Scores have been standardized by subtracting the mean score 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 score in Maths pre-university tests.

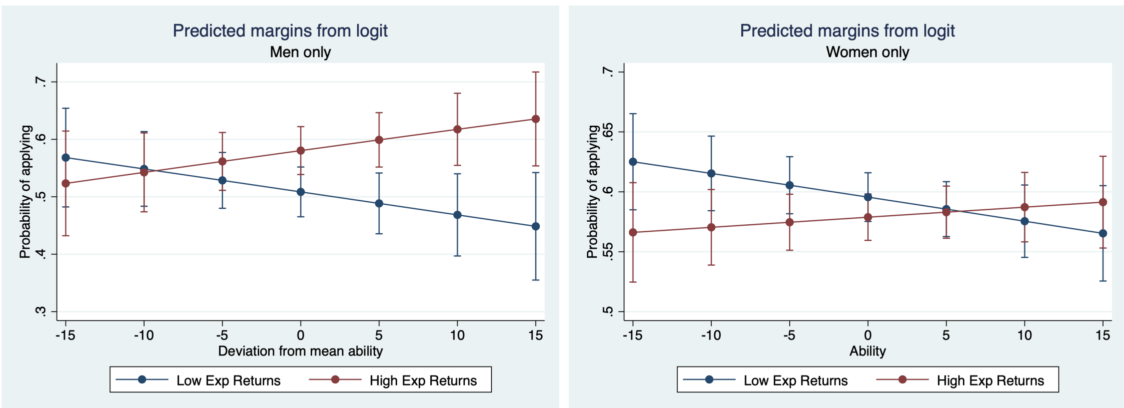
Table A.7. Women's on-the-job performance

	DV: First Semester Std. Scores	
	(1)	(2)
Male Photo	0.126^{+}	0.168*
	(0.082)	(0.091)
High Exp Returns	-0.031	-0.025
	(0.080)	(0.068)
Observations	955	955
R-squared	0.132	0.131
Basic Controls	Y	Y
Mean Dep Var	0.07	0.07

Clustered standard errors in parentheses (ind. level)
*** p<0.01, ** p<0.05, * p<0.1, $^{+}$ p<0.15

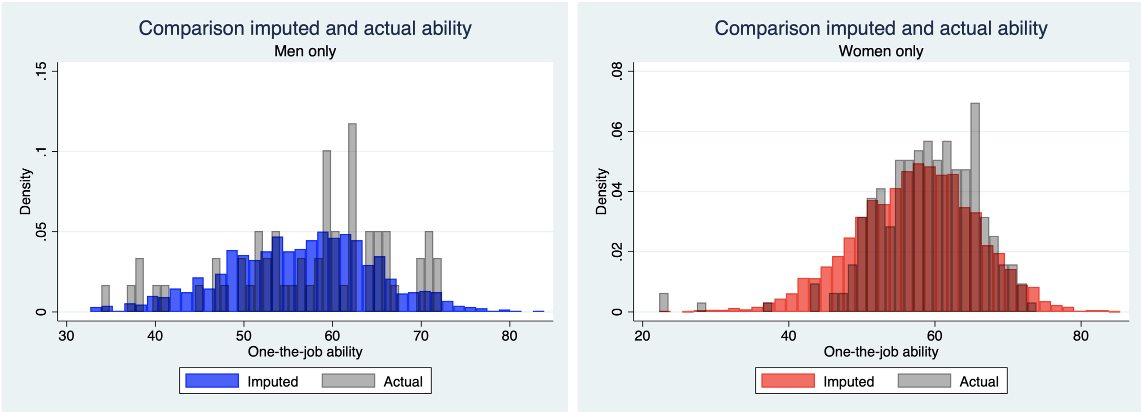
Note. OLS panel estimates for women only. The table reports results of two different regressions. The omitted category is the treatment group which received the female photograph and information of low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information of high returns to ability (specification (2) of Section 5.1). Column (2) introduces weights 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 score in Maths pre-university tests. Standard errors are clustered at the worker level.

Figure A.4. Predicted margins from logit by treatment



Note. The figure shows predictive margins from the logit discrete choice model. The graph on the left-hand side shows results for men and on the right-hand side for women. 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 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.

Figure A.5. 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). Imputed performance is calculated using a truncated linear regression 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.

Table A.8. 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	Y	Y	Y	Y
Outside Option Control	N	Y	N	Y
Mean Dep Var	0.835	0.835	0.798	0.798

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 of 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.9. 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	0.67	0.66	0.64	0.63

Robust standard errors in parentheses

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

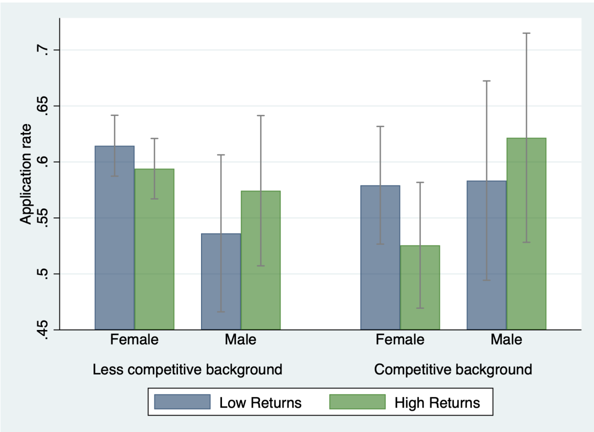
Note. OLS estimates. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment. All the regressions control for the basic set of controls X_i made of the following dummies: past application, access to early registration and non-white ethnicity. "Non hetero" is a dummy equal 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. "Age > med" is a dummy for age above median of the sample (of men and women separately).

Table A.10. 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
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**
			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 to a subsample of my experimental participants (N=633) to rate themselves in ten skills on a scale from 1 (max) 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).

Figure A.6. 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' occupational background. "Competitive background" is defined as having studied a male-dominated subject (e.g., engineering, business, math) in a top tier university in the U.K.. "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 of respondents 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 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).[102] 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 sample of current job applicants is meant to capture possible unobservability in characteristics of people interested in the particular job and/or organization. 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 and subject payment in detail.[103]

Amazon Turk photographs categorization. Respondents were Amazon Mechanical Turk workers who hadn't participated in any of the researchers' previous experiment conducted on the same platform and who have been granted the "Master" qualification on the website. The survey was conducted with the 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 of 161 answers, of which 39 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 20 cents.

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

[103] 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.[104] 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[105] 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) haven't 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 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 observables 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's 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

[104]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.

[105]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

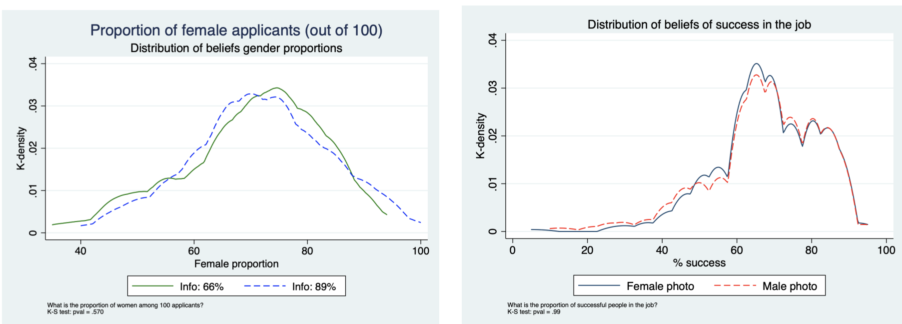
Note. Friendliness of the person in the picture was rated answering the question: "How does the person in the photograph appear to you?" on a 5-points scale. The variable "Friendliness" is a dummy equal to 1 if the person replied Friendly or Very Friendly and 0 otherwise. Work satisfaction was rated 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 man, whether the job is desirable for woman, whether they think that customers discriminate 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.

The main takeaway from Table B.2 is that respondents' beliefs about the quality of the pool of applicants and percentage of high performers in the job changes 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 of 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 that the two treatments are interacting, which would make hard the identification of the two separate channels.

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 of 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 didn't open the invitation-to-apply email, but unfortunately I don't 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 didn't look at the invitation-to-apply email is equally likely across experimental conditions, which should then only create an attenuation bias in the results.[106]

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 shouldn't 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 10).

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 3).

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.[107] 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 of being evidence-based and statistics are frequently reported on the organization website.

[106] 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).

[107] 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.

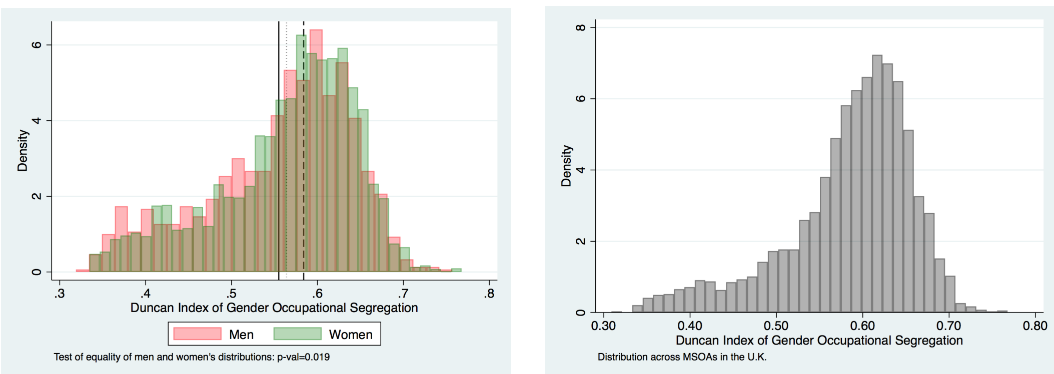
C Exposure to occupational gender segregation

C.1 Measures and methods

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 MSOA level. The sample is a 10% random sample from the 2011 Census, obtained through a special request to the National Statistical Office. MSOA stands for Medium Layer Super Output Areas. In 2011, the median MSOA in the UK comprised 188 8-digits postcodes, with a minimum of 89 postcodes to a maximum of 1033. There are 7201 MSOA in the UK in 2011. 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 in most of the cases the postcode of their parents' home, which is most likely 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 Table C.1. The U.K. average Duncan Index across MSOAs is 0.5839 and the average in my sample is 0.563. Table C.1 shows demographic characteristics by gender and exposure to high versus low gender segregation.

Figure C.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 2 are robust to assigning the Duncan index computed from the 2001 Census data to individuals older than 23 (60% of men's sample).

Table C.1. Demographics by exposure to occupational segregation

	Duncan < med			Duncan > med			Diff means
	Mean	SD	N	Mean	SD	N	P-val
				Men			
Non-white	0.31	0.46	498	0.25	0.43	498	0.03
Age	27.76	8.14	498	29.7	10.06	498	0
Married	0.15	0.36	487	0.24	0.43	491	0
Caring duties	0.13	0.33	498	0.19	0.39	498	0.01
Top university	0.29	0.45	498	0.21	0.41	498	0.01
First Grade	0.2	0.4	498	0.2	0.4	498	0.81
FTE	0.5	0.5	498	0.5	0.5	498	0.95
Outside Option	2.53	0.28	498	2.59	0.31	498	0
Aligned Subject	0.44	0.5	498	0.52	0.5	498	0.01
				Women			
Non-white	0.19	0.39	2167	0.36	0.48	2166	0
Age	26.71	8.17	2167	26.04	7.75	2166	0.01
Married	0.14	0.35	2137	0.1	0.3	2123	0
Caring duties	0.19	0.4	2167	0.14	0.34	2166	0
Top university	0.21	0.41	2167	0.29	0.45	2166	0
First Grade	0.19	0.39	2167	0.17	0.38	2166	0.16
FTE	0.44	0.5	2166	0.4	0.49	2166	0.02
Outside Option	2.42	0.24	2167	2.39	0.22	2166	0
Aligned Subject	0.73	0.44	2167	0.66	0.47	2166	0

Note. Differences in means between men (top panel) and women (bottom panel) who come from areas with occupational gender segregation above the median or below the median along demographic, educational and employment variables. The variable "caring duties" is a dummy equal to one if the respondent is a primary or secondary carer of children. I define top U.K. universities those belonging to the Russell Group. "First grade" is a dummy for whether the person got a first class in university. "Aligned Subject" is a dummy equal to one if the person studied a subject aligned with the job. "Outside option" is the expected log hourly-wage in the U.K. job market conditional on subject studied, gender, race, age, British nationality and marital status.

C.2 Occupational segregation, social norms and beliefs about gender

The validity of the proxies for α used in Section 6.3 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; Lordan and Pischke, 2016). Moreover, sociologists have been extensively studying the association between the former measure and gender attitudes (England, 1990). I present three data exercises to validate the proxy used.

First, I show that men who come from areas above the median of the Duncan Index display an implicit association bias 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 a Single-Target Implicit Association test (Greenwald et al., 1998).[108] I designed an ad-hoc test to measure the extent to which respondents automatically associate social work with women.[109]

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 (Greenwald et al., 2003). 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 positive the d-score the stronger is the association between the two concepts.[110]

Figure C.2 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

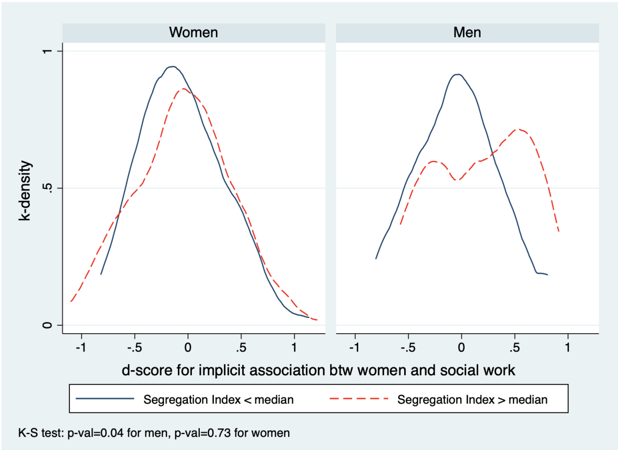
[108] Response rate was 12.5% for the main survey and 6% to the IAT (604 and 300 respondents respectively).

[109] Many studies in economics have used the IAT as a predictive measure of employers' discrimination (Bertrand et al., 2005; Reuben et al., 2014; Glover et al., 2017) or sensitivity to negative stereotypes (Cvencek et al., 2011; Nosek et al., 2002; Kiefer and Sekaquaptewa, 2007; Carlana, 2018). For a recent review, see Bertrand and Duflo (2017).

[110] The order of the two types of blocks was randomized at the individual level.

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), teachers (Carlana, 2018) and employers (Ruben et al., 2014). I provide evidence that labour market conditions correlate with implicit biases held by the minority, but I do not find evidence for behavioural consequences.

Figure C.2. 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").

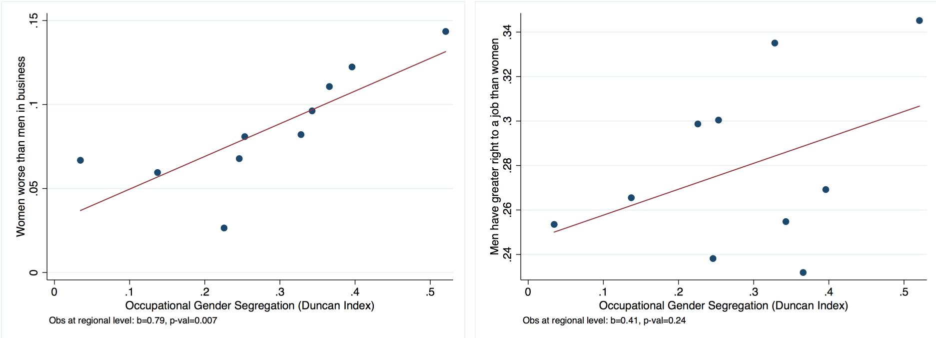
In Figure C.3 I show that U.K. regions with high gender segregation levels display more traditional norms related to women's employment. In the two scatter plots of Figure C.3, the x-axis shows the proportion of local authorities in a certain region that have a Duncan Index in the top quartile of the 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.2 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:

- 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)
- 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 Y?

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

Figure C.3. 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.

variable in Table C.2. 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 14 years old. The regression controls for ethnicity, survey sample and the level of beliefs elicited in the first question mentioned above. We can immediately see 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.2. 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 DV	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 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 Mincerian regression of the log-hourly wage on a set of observables which are available both in the LFS and my experimental dataset.[111] 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 describe the exercise in detail in the next subsection.

I interpret this measure as the individual outside option component w^o. While providing a useful measure of the candidates' opportunities in the labour market at the time of application, the 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.[112] Table D.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 graduates 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.

Table D.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
Sciencific 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. Column "Diff (1)-(2)" contains the difference in the proportions of women and men that have the characteristic of the corresponding row. "FTE in Public Sector" is an indicator variable for working in the government and includes jobs in healthcare.

[111] I used the following set of dummies: university subject (16 categories), age, age squared, British nationality, gender, marital status, non-white ethnicity, first grade in university.

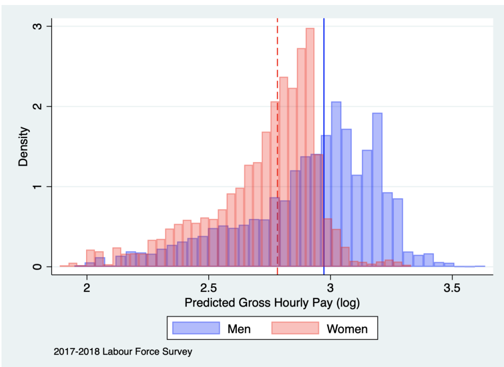
[112] For instance, the LFS data do not contain the exact university attended by the respondents.

Methodology

I use data from the eight quarters of the 2017 and 2018 Labour Force Survey in the UK.[113] I limit the sample to men and women between 16 and 64 years old. To match the eligibility criteria of my experimental sample, I exclude from the sample people who don't have at least a bachelor degree or, if students, who are not currently studying towards a bachelor degree or higher university title. I then estimate a Mincerian regression of the hourly pay of people in employment on the following series of dummies: university subject (JACS3 macro areas), being married, being of non-white ethnicity, being a man, being born in the UK, age and age squared, having obtained a first grade in university.

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). As the distribution looks log-normal, 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 extract the coefficients of the estimation and apply them to the same variables in my experimental data, in order to construct a predicted individual outside-option. I decided not to control for the fulltime employment status because the coefficient would bias upward the estimated outside option of people in fulltime employment as compared to both students in my sample. We don't know whether the people who are students in my sample will decide to become full-time workers or not; thus the estimated outside option for students would be biased downward if they will become full-time employees. Figure D.1 shows the distribution of the computed outside option by gender. Table D.2 shows the coefficients of the Mincerian regression on the LFS data. The omitted category are non-married white women who studied Arts.

Figure D.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.

[113] For more information on the Labour Force Survey, see the LFS website.

Table D.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.

E Performance on the job: distributional effects

In this section, I look at the impacts of the treatments on the quality of hired workers by measuring changes in the conditional quantiles of workers' 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 E.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

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 of high returns to ability treatment. The outcome variable is the index of desirable qualifications 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.

Table E.2. On-the-job performance: quantile regressions

DV: First Semester Std. Scores					
	(1)	(2)	(3)	(4)	(5)
	Quantile				
	10	30	50	70	90
	Women only				
Male Photo	0.118	0.066	0.111*	0.105*	0.028
	(0.149)	(0.061)	(0.067)	(0.054)	(0.060)
High Exp Returns	-0.118	-0.027	-0.037	-0.011	-0.056
	(0.133)	(0.068)	(0.072)	(0.056)	(0.054)
Observations	955	955	955	955	955
R-squared	0.120	0.097	0.093	0.113	0.085
	Men only				
Male Photo	-0.029	0.040	0.044	-0.008	0.026
	(0.397)	(0.254)	(0.136)	(0.112)	(0.200)
High Exp Returns	0.499	0.198	0.132	0.083	-0.053
	(0.364)	(0.209)	(0.117)	(0.108)	(0.111)
Observations	215	215	215	215	215
R-squared	0.145	0.195	0.186	0.171	0.059
Exam FE	Yes	Yes	Yes	Yes	Yes
Control for Quality	Yes	Yes	Yes	Yes	Yes

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

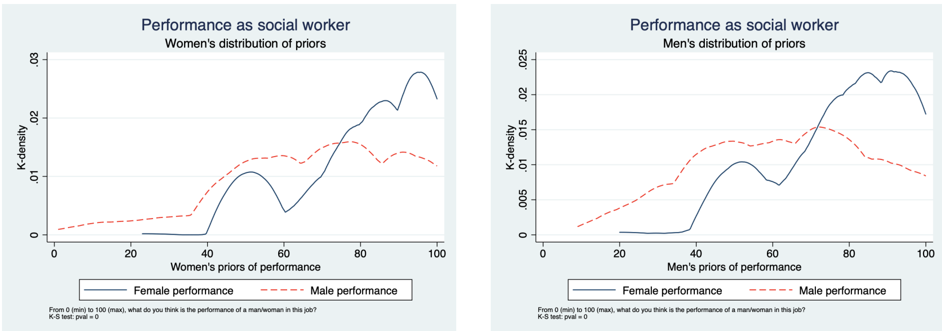
Note. Quantile regression with panel data. Estimations are for women in the top panel and for men in the bottom panel. The omitted category is the treatment group that received the female photograph and the low returns information. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment. The regressor "High Exp Returns" is a dummy equal to one for information of high returns to ability treatment. The outcome variable is the standardized score obtained in the five assessments required within the first semester on the job. 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, being from a top tier university and score in Maths pre-university tests. Standard errors are clustered at the worker level.

F Appendix to theoretical framework

F.1 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 F.1 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 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 $\overline{\sigma}_M^2 > \overline{\sigma}_W^2$.

Figure F.1. 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 F.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 (or unbiased expectations of) a_i seems appropriate.

Table F.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.

F.2 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.

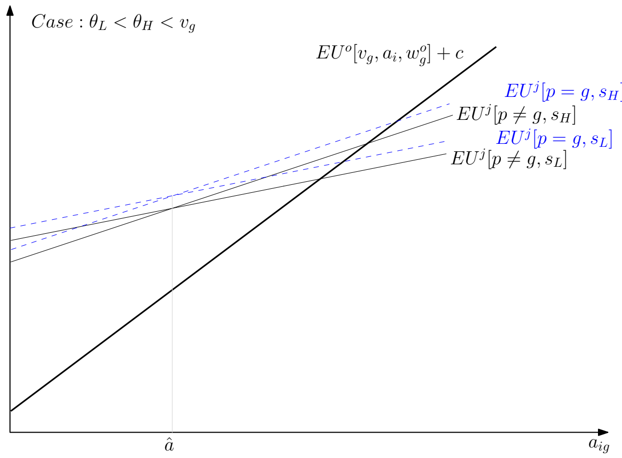
Result 3. ***Interaction between gender shares and expectations***

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 F.2 provides the graphical intuition for Result 3 for the case $U^{j'}(a_i) < U^{o'}(a_i)$.

Figure F.2. 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 of 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 of 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.

F.3 Adding stereotypes to the model

I assumed so far that photographs have no effect on information interpretation. Yet, in the experiment as well as in the real world, the frame or context where information is conveyed can affect learning. The photographs manipulation might interfere with people's updating of expected returns to ability on the job.[114] Recent work on beliefs in gendered domains (Bordalo et al., 2016, 2019; Coffman et al.,

[114] There is rich experimental evidence on people's "mental gaps" in information gathering and processing (Handel and Schwartzstein, 2018), such as neglecting important information components (Schwartzstein, 2014) or overweighting salient features (Bordalo et al., 2013, 2017). Different pictures could differentially catch people's attention, making them more or less attentive to information (see Section 10). Or there could be a small probability that potential applicants attribute the aggregate statistics to the gender group portrayed in the picture. Such an effect - a sort of group attribution error (Fiske and Taylor, 1991) - can arise from limited attention or rational uncertainty, if people are not sure if the statistics received refers to everyone. This hypothesis can be formalized as a higher signal precision σ_s^2 when $p = g$, with consequently larger $d\theta_g$. I don't find evidence for this effect.

2019), as well as the literature on confidence by gender and task content (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 with women, become more optimistic about own performance as female advantage increases. My results are 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. One possibility is that the male photograph makes women revise their gender advantage in the job. I follow this line of thought to 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).[115] 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. Let's take an extreme case for the sake of explanation. In the male photograph treatment $\hat{a}$ is greater than in the female photograph treatment. If this difference is big enough, it can lead to a situation in which condition B ($a_i^* > \hat{a}$) is satisfied in the female photograph treatment and violated in the male photograph treatment. This implies, in turn, that the difference in application rates between receiving information of high or low returns is positive conditional on a female photograph and negative conditional on a male photograph. Thus in this model application rates are not necessarily the lowest in treatment $(p \neq g, s = s_L)$ but can the lowest in treatment $(p \neq g, s = s_H)$.

F.4 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 $[a_1, a_2]$

[115] 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 F.1. 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 hypothesised effect.

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

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

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.

□

G Dynamics

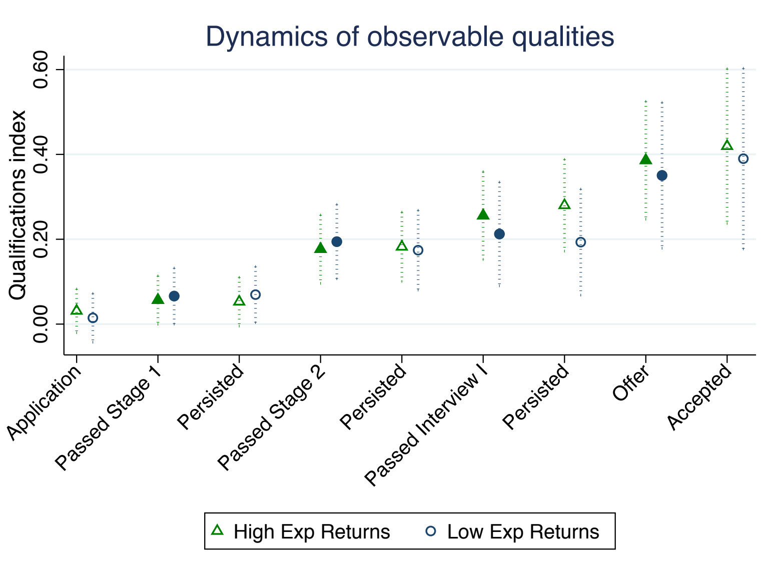
I provided evidence on the cumulative treatment effects over selection stages, bundling together the effect on application submission and withdrawals across stages. This section presents evidence on the dynamics of treatment effects across the four stages of the selection process: stage (I), stage (II), interview (I) and interview (II). Figure G.1 shows the dynamics of individual decisions to remain in the process. For instance, it shows that 91% of men in the high returns treatment decided to show-up to interview I (conditional on having succeeded in Stage 2). There are two take-aways. First, the information treatment affect men's decision making over time, not only in the very first stage. Secondly, the impact of the information treatments on individual decisions is greatest - and in the same direction - in the two most time-consuming stages: application submission in stage I (which takes between four and six hours) and interview II (which is half day long). The dynamics of treatment effects for women are concentrated in the first stage instead.

Figure G.1. Dynamics: stayers over the hiring process (men only)

Note. The figure shows the proportion of men who decided to go on to the next stage, for each of the four stages in the selection process. The blue solid line is for the low expected returns treatment and red dashed line for the high expected returns treatment. For instance, the graph shows that 91% of men in the high expected returns treatment decided to show-up to Interview I (conditional on having succeeded in Stage 2).

Does quality differ across stages of the hiring process? Figure G.2 uses the index of quality computed by averaging 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 tests. This is the same set of variables used for the index reported in Figure A.2 and in Table E.1. To define cognitive skills, I use the employment history reported by each applicant in the application form. Each applicant can list up to two previous employers, specifying the role covered, the level (e.g., junior, senior with or without management responsibilities) and the main duties. I coded the most recent 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. For each person, the measures of cognitive and manual skills should thus be interpreted as the average level of cognitive and manual skills acquired at work.

Figure G.2. Dynamics: qualifications over the hiring process



The figure shows men's average proportion of desirable qualities in all the stages of the hiring process. "Desirable qualities" are measure with an index between 0 and 1 that includes the following variables: having a first grade in university, being from a top tier university, having studied a subject aligned with the job, having volunteered frequently in the past, having cognitive skills above the median and having manual skills above the median. Hollow symbols refer to the four hiring stages. Full symbols refer to intermediate stages in which candidates can decide whether to persist in the process. Blue line is for the high % information treatment and red for the low % information treatment.

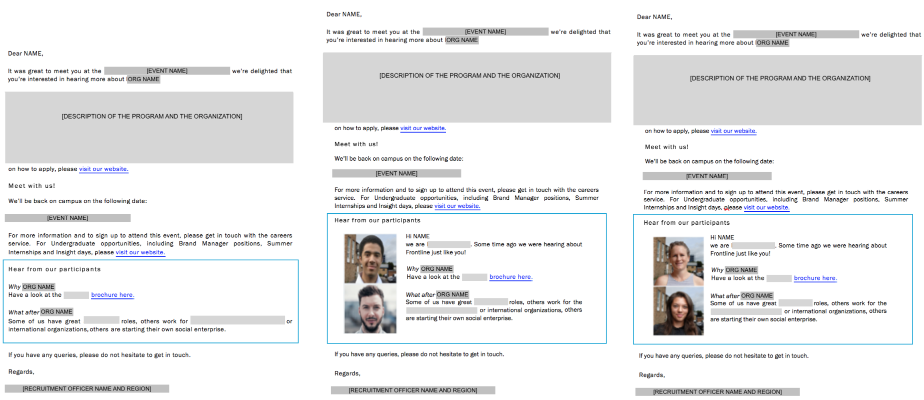
Figure G.2 shows the average proportion of "desirable" observable qualifications that men have in each stage of the hiring process, by information treatment. Full symbols refer to the four stages which involve screening by the employer, as in the previous Figure. Hollow symbols refer to intermediate stages in which candidates can decide whether to persist in the process. In these stages it's only the applicants' decision whether to persist in the hiring process or withdraw. Figure G.2 shows that the big difference in quality between the two treatments appears after Interview 1, as a result of candidates' decision to stay in the process. This suggests that providing information on high returns to ability is not only effective to attract more applicants, but also to keep the best ones in the selection pipeline.

H Do gender shares matter for a wider pool of students?

I partly 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.[116] Mailing lists were collected in 75% of the total number of events run by the organization.[117]

Figure H.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.[118] 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 H.1. Assignment to treatment was stratified by university, event and gender.

Each email template 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.[119] Online behaviour is measured using standard metrics

[116] RO's performance evaluation does not depend on the number of email addresses collected at university events.

[117] 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.

[118] Given that sign-up takes approximately 30 seconds, thus experimental subjects likely have different levels of interest in the job, but at least a minimum level of attention to it.

[119] To be trackable, unique links at the university-event-gender-treatment level were created before the randomization

recorded by the Google Analytics service installed on the organization website. 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, for a total of 2877 unique participants (630 men).[120] Table H.1 presents summary statistics of the sample in Experiment 1 and balance checks. 78% of participants are last year students or graduates, while the remaining proportion are first or second year students; 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 H.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 as compared to the control group across all events. The number of clicks almost doubles (Figure H.2). Despite this first stage, behaviour does not translate into more applications. Table H.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:

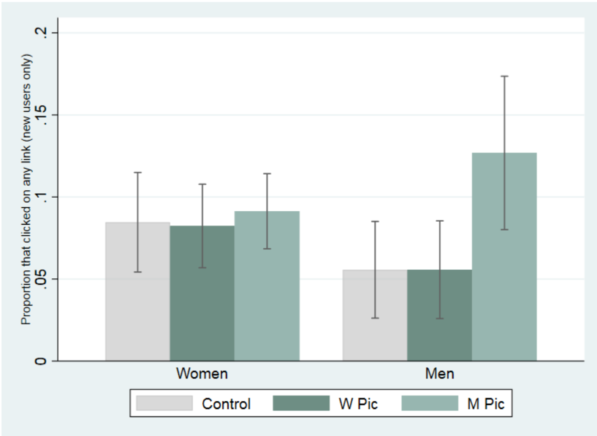
$$y_{geu} = c + \beta_1 MPic_{geu} + \beta_2 WPic_{geu} + X'_{eu}\beta_3 + \delta_u + \epsilon_{geu}$$

The regression includes 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 H.2 shows that the treatment per sé doesn't increase applications, which reinforces the external validity of the null effect of the male photograph.

by adding an alpha numeric snippet to the website url.

[120] Mean participation covers substantial variation between event types: stands at career fairs had an average number of attendees around 44 compared to an average of 16 for presentations and panel events.

Figure H.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 H.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 DV	0.0821	0.167

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.

I An additional exercise on overconfidence

In this section I further explore whether the effect of the information treatment can be explained by overconfidence and potentially gender differences in it. I use the survey questions defined at the end of Section C.2 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 I.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.[121] 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. Repeating the same exercise on women shows that information of high returns to ability discourages applications by women with below median confidence in men's performance in female-dominated jobs. This seems consistent with priors' precision being correlated with lower confidence also in own ability.[122]

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

	DV: Applied and never DO = 1			
	(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 DV Pure Control	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 of low expected returns to ability.

[121] 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.

[122] The negative impact of the male photograph on women's applications is identical across levels of predicted confidence in both men's and women's performance.