The London School of Economics and Political Science

Essays in Development, Gender and Personnel Economics

Alexia Delfino

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy
London, September 2019

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of approximately 40,000 words (including footnotes but excluding bibliography and appendices).

London, *30th September 2019*

Alexia Delfino

Statement of Conjoint Work

I confirm that Chapter 2, "Rule of Law and Female Entrepreneurship", was jointly co-authored with Professor Nava Ashraf (LSE) and Professor Edward Glaeser (Harvard). This statement is to confirm that I contributed 33 percent of this work.

London, *30th September 2019*

Alexia Delfino

Abstract

The first two chapters of this thesis provide insights into the determinants of occupational gender segregation in both developed and developing countries. The third chapter of this thesis goes beyond gender to understand how different aspects of individual identity, namely values, affect selection and performance in the workplace.

Across the developed world, traditionally female-dominated sectors are growing and traditionally male-dominated sectors are shrinking. And yet, sectorial male shares are not changing accordingly. Why don't men enter female-dominated occupations? In the first chapter, 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.

The second chapter deals with gender segregation in the growing urban marketplaces of the developing world. Commerce between strangers requires trust, but trust is difficult when one group consistently fears expropriation by another. If men have a comparative advantage at violence and there is little rule-of-law, then unequal bargaining power can lead women to segregate into low-return industries and avoid entrepreneurship altogether. In this paper, we present a model of female entrepreneurship and rule of law that predicts that women will only start businesses when they have both formal legal protection and informal bargaining power. The model's predictions are supported both in cross-national data and with a new census of Zambian manufacturers. In Zambia, female entrepreneurs collaborate less, learn less from fellow entrepreneurs, earn less and segregate into industries with more women, but gender differences are ameliorated when women have access to adjudicating institutions, like Market Chiefs and a Small Claims Court. We experimentally induce variation in local institutional quality in an adapted trust game, and find that this also reduces the gender gap in trust and economic activity.

The third chapter of this thesis studies the primitives of corporate culture: employees' values. It is well known that shared values can mitigate the adverse consequences of incomplete contracts and reduce coordination costs. Misalignment in values - either actual or perceived by the employees - could make such inefficiencies worse, reducing productivity and creating resistance to change. We provide evidence by means of a survey that measures the perceived and actual value misalignment between employees, their colleagues, and top management in a multinational bank. The data, which covers 30,000 employees across 55 countries, reveals that values dissonance is negatively correlated with individual and team level performance, as well as with self-reported trust in the bank's executives and intent to stay. Within countries, we show how bankers' values compare with the ones listed by World Value Survey respondents. Employees whose values are further from common citizens perform better, but this is mainly explained by their higher position on the career ladder. We conclude by showing how aggregate shocks to societal perceptions of banking, such as the 2008 financial crisis, might shape organizational culture.

Acknowledgements

I would like to first express my deep gratitude to my supervisors and advisors. I feel blessed for having had the opportunity to work with all of them. I want to thank Nava Ashraf for believing in me. She shaped my development as a researcher from many different perspectives and taught me the value of being bold in thinking differently. She has been an exceptional example of academic passion combined with rigor and deep insights into human nature. I am grateful to Oriana Bandiera for constantly supporting and encouraging me. She set for me an unachievable standard of efficiency and smart economic thinking combined with kindness and care. I am thankful to both of them, Nava and Oriana, for guiding me through challenging partnerships and sharing unexpected early meetings in Canary Wharf. I am indebted to Matt Levy for always being supportive and willing to dedicate time hearing my confused thoughts. He reminded me to be more concrete and proceed step-by-step, helping me not only with smart insights into all my projects, but also with emotional support. I am grateful to Erik Eyster for our conversations and for pushing me to figure out how a simple set of variable should capture important mind processes. Working with Edward Glaeser has been extraordinary and the number of things I learnt from him is uncountable. I thank Robin Burgess for helping me see the big-picture of my work and his support in training my confidence.

I have been lucky to benefit from the advice and insights of many other professors at the LSE: Gharad Bryan, Maitreesh Ghatak, Tim Besley, Greg Fischer, Jordi Blanes i Vidal, Guy Michaels, Alan Manning, Daniel Sturm. I am indebted to Luigi Marengo and Alessandro Nuvolari for their mentorship over many years.

I thank Mark Wilbor, Jane Dickson, Angela Swain, Rhoda Frith and Lubala Chibwe for their help in a variety of requests. I acknowledge the financial support of the Economic and Social Research Council (ESRC), STICERD and Fondazione Einaudi. For the second chapter, I acknowledge generous financial support by IGC and PEDL.

Many great people enriched my journey through the PhD and reminded me that the ultimate purpose of what we are doing should go beyond the walls of some fancy university rooms. Kami Lamakan has been a constant source of inspiration and motivation. I thank Katie Mond and Katie Purser for their relentless enthusiasm about our joint work and all people in the partner organization that dedicated time and effort to make it happen. For the second chapter, I thank the IPA teams that worked on data collection, the support of the IGC country economists and the sleepless nights of Calvin Chiu, Kate Laffan and Pascale Bourquin. I thank Neil Campbell for providing the data for the third chapter and for being so helpful in our collaboration.

I thank the invaluable research assistance of Cristhian Acosta, Alex Blums, Andres Cordoba, Pedro Cabra, Nick Swanson, Stefan Faridani, Miguel Fajardo, Beatrice Montano and Edward Davenport.

I feel honoured to have shared this ride with many brilliant people: Svetlana Chekmasova, Clement Minaudier, Niclas Moneke, Giuseppe Rossitti, Giulia Bovini, Vincenzo Scrutinio, Michel Azulai, Dana Kassem, Arthur Seibold, Albert Brue, Shan Aman-Rana, Celine Zipfel, Claudio Schilter, Xuezhu Shi and Pedro Alves, among others. I thank Giulia Giupponi for being the best "buddy" I could wish for and for always been there for me. I thank Francesco Sannino for making me laugh with his gossips and being an unexpected source of positive motivation. Gianpaolo Caramellino has been a calming and generous presence, always happy to give advice. Irene Lauro's energy provided a unique blast of lightness in this journey. I thank Kim Sarnoff for our deep conversations in front of rubbish bins. Nicola Limodio and Matteo Benetton have been my big brothers. I thank Matteo for always been there for me, in the best and in the worst moments, ready to go for a walk or a drink, or simply to stay at the kitchen table with our laptops sharing silly and serious stories. Nicola is an infusion of energy, perseverance and hard work that lifted me up when I needed the most.

I thank members of "Gruppone", my friends from Pisa. No matter the distance between us, we keep cheering for each other's success and look forward to celebrating together. I thank Chiara for our telepathy and true friendship, for always being there to help me take the right decisions and for the cruise we keep postponing.

I could not have accomplished any of this without the constant support of my family and no word could ever be enough to express my gratitude to them. I thank my nephews for bringing sunshine to my stay in London. Mattia helped me settle in my first flat, sharing a cold room with no blankets and quite a few spiders. Long walks at the Zoo with Davide's story telling inspired the start of the first chapter. I hope this thesis will contribute to make them proud of their auntie. I thank my sister Manuela for always being ready to rescue me, not matter how far I am, how little she sleeps and how busy she is. She is my Guardian angel and I feel extremely lucky for her unconditional love. I thank Fabio for introducing me to experimental research and for being my first mentor and supporter, willing to give priceless advice at any time and on any topic. I thank my parents for their love, unconditional support and evening calls, which made me feel closer to home. The memory of their hard-work and resilience as migrants never abandoned me in these years. I thank my mum for always encouraging me to keep going and taking care of many practical issues for me so that I could focus on my work. I thank my father for never doubting about my talent and transmitting his passion about teaching with both words and acts. Long nights at the office allowed me to meet Miguel and this work was born between our Piccadilly-line midnight carriages and Ryanair flights. I thank him for always standing by my side, being blind about my flaws but long-sighted about my qualities. His past and present story is an example of fearless motivation that I learn everyday from.

To Miguel, the most unexpected
and precious gift of this journey.

Contents

List of Figures

List of Tables

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

1.1 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 help workers in declining industries move 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 selection.

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 the behaviour of participants 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]

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.

[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 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).

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-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

[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 1.3 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).

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 does 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.

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

[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.

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

[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 1.5.

[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).

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 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]

1.2 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

[16] Along these 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.

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 represent more than 75% of social workers across the developed world, as shown in Figure 1.1 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 1.A.1, Bureau of Labor Statistics 2019).

Figure 1.2 illustrates the timeline of the organization's 2017 nationwide recruitment. The experiment happened between September and November, which is the application period.[22] 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

[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.

[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] The experiment is registered in the AEA RCT Registry with ID AEARCTR-0002351 and was approved by the LSE Ethics Committee in August 2017.

is around ten weeks. If a person was hired and accepted the job, actual work in local authorities started in July 2018.

1.3 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.[23] I introduce exogenous variation in the content of the invitation-to-apply email along two dimensions: perceived gender shares and expected returns to ability.[24] The two experimental conditions were cross-randomized in a fully nested design, leading to a total of four treatment emails. Participants could also be randomly assigned to receive a fifth "pure control" email containing no manipulation, which I used to compare the treatments with business-as-usual for the organization. Randomization was at the individual level, with stratification by gender (man/woman) and ethnicity (white/non-white). The experiment was double-blinded: participants were not aware that the invitation-to-apply email was part of a research study and recruiters were not aware of candidates' treatment assignment. This design limits experimental biases that arise from candidates' knowledge of being part in a research study and prevents recruiters' assessment of candidates from being influenced by their treatment.[25] 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.[26] I use my theoretical framework to show that these two interpretations are observationally equivalent

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

[24] 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.

[25] 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.

[26] 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).

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).[27]

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 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.[28] Other issues might arise if there is a systematic correlation between portrayed workers' characteristics and their gender. I discuss these concerns in Appendix 1.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.[29] Moreover, the effect of individualized information on beliefs depends the level of people's priors, which was unobservable to me.

[27] 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.

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

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

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 1.3):

> 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).[30]

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.[31] 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 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.[32]

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 1.10.

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

[30] 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.

[31] 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.

[32] 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.

in generating output for the employer.[33] 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).[34] 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).[35]

Figure 1.3 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%).

1.3.1 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.[36]

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.[37] Both samples were incentivized for participation and the survey had an average completion time of 15 minutes. Appendix 1.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,

[33] 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.

[34] 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.

[35] 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.

[36] 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).

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

the survey elicited beliefs on a variety of characteristics of the job and the pool of applicants. Figure 1.4 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 1.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 1.5 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.[38] The right panel of Figure 1.5 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.

[38] 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.

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

1.4 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.

1.4.1 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 $a_i \sim U[.]$.[39] 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 $\overline{w}_g = w_g^o - w$.[40]

[39]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).

[40]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.

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. [41] In this view, θ_g can be interpreted as the believed marginal product that a person of gender g with ability a_i achieves in the job and which determines either monetary or non-monetary gains. The parameter $\hat{a}_g$ is the level of ability which is not affected by changes in marginal returns to ability, while ability levels above (below) $\hat{a}_g$ get higher utility from higher (lower) θ_g.[42]

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 $s_g \sim N(\bar{s}_g, \sigma^2_{s_g})$ with $\bar{s}_W > 0.5$ 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.[43]

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: $\overline{\sigma}^2_M \geq \overline{\sigma}^2_W$.

[41] 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.

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

[43] 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, $-\alpha_{iM}$ can be the cost of social punishment for selecting a female job and s_g is the likelihood that the punishment will be enforced.

The combination of assumptions 1 and 2 is equivalent to assuming risk aversion in the utility function and keeping only the assumption of asymmetric uncertainty.[44] Appendix 1.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.[45] 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.

1.4.2 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.[46] The employer includes two pieces of information: the photograph 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 $s \in \{s_L, s_H\}$.[47] 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.[48]

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

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

[45] This evidence come from the auxiliary online experiments described in Appendix 1.B.

[46] Workers' diverse composition might positively affect output through different channels, for instance through skills complementarities (Lazear, 1998), better matching between clients and employees (Hoogendoorn and Van Praag, 2012) or organizational reputation (Erhardt et al., 2003; Carter et al., 2003; Flabbi et al., 2019).

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

[48] 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.

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 1.3.1).

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 $\theta_g^{'}$ 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:

$$\Delta\theta_g = (\theta_g|s_H) - (\theta_g|s_L) = \frac{\overline{\sigma}_g^2}{\overline{\sigma}_g^2 + \sigma_s^2} \cdot (s_H - s_L) \tag{1.1}$$

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

1.4.3 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 $U^j(a_g^*) - c = U^o(a_g^*)$.[49] 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.

[49] See Appendix 1.F for the formal proof.

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 1.6 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.[50] Define $B = \alpha_i s_g - \overline{w}_g - c - v_g \hat{a}$.

Result 2. ***The effect of a shock to expected returns to ability***
If $B > 0$: when $s = s_H$, the pool of applicants N_g is larger than when $s = s_L$.
If $U^{j'}(a_i) < U^{o'}(a_i)$ and $B > 0$: when $s = s_H$, 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.[51] Figure 1.7 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

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

[51] 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 $s = s_L$.

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.[52] 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 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 $\bar{\sigma}_g^2$. The implication is that, ceteris paribus, updating will be stronger for men than women because of their higher $\bar{\sigma}_g^2$.

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.

1.5 Sample, balance and empirical strategy

The experimental sample consists of 5417 candidates, of whom 1013 are men. Table 1.1 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.[53] 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.[54] 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.

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

[53] 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.

[54] A care leaver is a person who has been looked after by a local authority for at least 13 weeks since the age of 14. In 2012, the total number of care leavers represented 0.12% of the total UK population between 16 and 25, while they make up 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).

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 1.1 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 1.1 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 1.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.[55]

1.5.1 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.[56] I perform

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

[56]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.

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.

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

$$y_i = c + \beta_1 Pic_i^M + \beta_2 Returns_i^H + X_i'\lambda + \epsilon_i \tag{1.2}$$

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.[57]

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.[58] 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 1.3).

In model (1.2), 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

[57] 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).

[58] 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).

combine the two manipulation to study whether there is an effect of their interaction.[59] I use the following specification:

$$y_{ig} = c + \delta_1 \cdot Pic_i^g \cdot Returns_i^H + \delta_2 \cdot Pic_i^{-g} \cdot Returns_i^L + \delta_3 \cdot Pic_i^{-g} \cdot Returns_i^H + X_i'\lambda + \epsilon_i \tag{1.3}$$

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 (1.3) uses the email that combines the same-gender picture with low-returns information as omitted category.[60] Model (1.3) tests three null hypotheses: $\delta_j = 0$, with $j \in \{1, 2, 3\}$.

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).[61] 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.

1.6 Results: men's entry

1.6.1 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.[62] 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 1.2). 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).[63]

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

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

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

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

[62] Figure 1.8a shows application rates across treatment groups from the raw data.

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

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 1.H). This extends the external validity of the null result of Table 1.2 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.[64] 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 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.[65]

1.6.2 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 1.2. 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.

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

[65] I discuss this point further in Section 1.10.

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 1.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)).[66]

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.

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 and 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).[67] However, the magnitude seems relatively large when compared to the effect of doubling posted wages in Abebe et al. (2019), who get an increase

[66] 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 1.9.

[67] 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.

in application rates of 18%, and Dal Bó et al. (2013), who show an increase of 26% for a 33% higher wage. 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 1.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.[68]

1.6.3 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).[69] 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 persistence of long-term statistical discrimination.[70] 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.[71] Using a bridge, I merged the index with my experimental data through

[68] Manipulation checks discussed in Section 1.3.1 also gave reassurance of this assumption.

[69] 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).

[70] 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".

[71] 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).

the subjects' secondary school postcode and, when missing, home postcode.[72] 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 1.3 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 1.3 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 the interpretation of Table 1.3 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.[73] Columns (3) and (4) of Table 1.3 repeat the same exercise using a different index: the average share of men working in female-dominated occupations in the local labour market.

Appendix 1.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.

1.6.4 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. 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

[72] 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 1.3 using only the subset of people with data on school location, but power drops.

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

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 1.4, 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 1.4 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 1.4 repeat specification (1.2) 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 1.D contains a detailed summary of the methodology used.

The evidence reported in Columns (3), (4) and (5) of Table 1.4 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 1.7).

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

[74] This index of wage dispersion is a function of the endogenous choice of university subject made by the candidates. Thus Table 1.4 could capture heterogeneous treatment effects due to other unobservable differences between candidates who chose the same university subject. As a robustness check, Table 1.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).

ability should also be more qualified. I provide evidence consistent with this prediction in the next section.

1.7 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.

1.7.1 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 1.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 1.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 1.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 1.2, 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 1.5, which shows the coefficients of the following regression:

[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.

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

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 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]

1.7.2 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]

[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

Figure 1.9a 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 repeated measures for each person are available.

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 1.9b 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

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 1.G.

[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).

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:

$$score_{ia} = \alpha + \beta_1 Pic_i^M + \beta_2 Returns_i^H + X_i'\lambda + \epsilon_{ia} \tag{1.4}$$

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 (1.2), 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 1.6 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 shown in Column (2) of Table 1.6, which adds weights for the difficulty of the local community where a worker is allocated to.[84] Appendix 1.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 1.7 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.

[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.

[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.

1.7.3 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 1.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 1.2 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 1.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]

1.8 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 1.8 shows that there are 7.5% fewer women's applications in

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

[87] Figure 1.8b shows application rates across treatment groups from the raw data.

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 1.8. 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 1.10 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 1.A.7 and 1.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 25^{th} and 75^{th}, see Table 1.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 1.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 1.8 shows that the point estimate on the $Returns_i^H$ dummy is -0.015 and far from being statistically significant. Column (1) of Table 1.A.3 confirms that the two genders react differently to the expected returns 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.

[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 1.A.3 shows that the effect of the male photograph on female performance does not show a decreasing trend over time.

1.8.1 Do women care about the workplace gender composition?

Women's reaction to the photograph manipulation could suggest that women value co-workers gender more than men do. 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 1.8). Secondly, the interaction between the two manipulations is important. The dashed lines of the right panel of Figure 1.11 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 1.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.

[89] See also the discussion in Section 1.10.

1.9 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: $Pr(apply = 1) = Pr(U^j(\alpha_i, s_g, \theta_g, a_i, \hat{a}) + \xi_j > U^o(v_g, a_i, \overline{w}_g) + \xi_o)$, 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 1.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 1.12 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.

[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 $F(z) = \frac{e^z}{(1+e^z)}$.

The right graph of Figure 1.12 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 1.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. This is consistent with the interpretation of the information treatment as a change in the slope of expected utility with respect to job-specific ability.

1.10 Alternative mechanisms

1.10.1 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 1.B) shows 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.

1.10.2 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

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

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 1.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 1.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.

1.10.3 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 1.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

[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.

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]

1.10.4 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 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

[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.

the number of skills rated above the sample mean.[97] Table 1.A.10 shows that men in my sample tend to be less overconfident than women, especially in job-specific skills. Appendix 1.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 1.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]

1.11 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

[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.

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 1.9 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 1.9 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.

1.12 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 1.10).

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

[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).

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.

1.13 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*. 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. B., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4), 1753-1794.

Bordalo, P., Coffman, K. B., 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.

Breda, T., Grenet , J., Monnet, M. & Van Effenterre, C. (2018). *Do female role models reduce the gender gap in science? Evidence from classroom interventions in French high schools*. PSE Working Paper 2018-06.

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, K. B., 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., 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.

Conlon, J. J., Pilossoph, L., Wiswall, M. & Zafar, B. (2018). Labor Market Search with Imperfect Information and Learning. NBER Working Paper w24988. National Bureau of Economic Research.

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.

Porter, C., & Serra, D. (2017). *Gender Differences in the Choice of Major: the Importance of Female Role Models*. Southern Methodist University Department of Economics Working Paper 1705.

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.

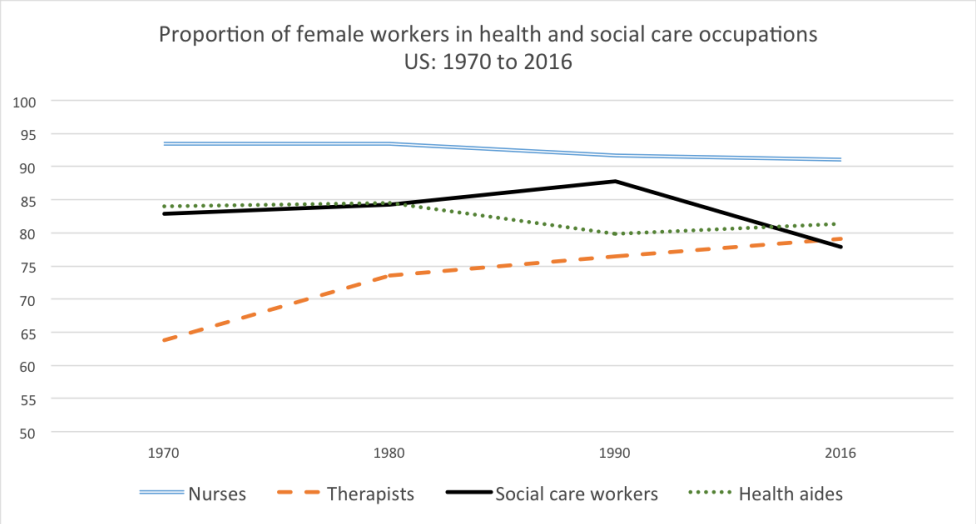
Wright, R., Kircher, P., Julîen, B., & Guerrieri, V. (2017). *Directed search: A guided tour*. NBER Working Paper w23884. National Bureau of Economic Research.

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.

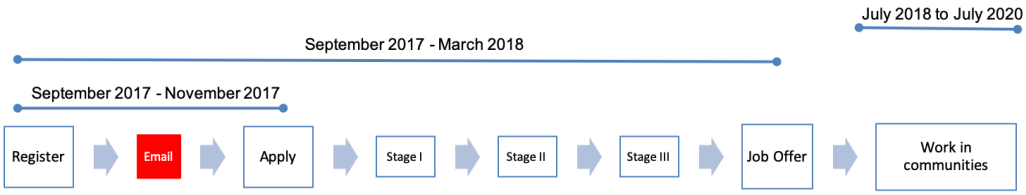
1.14 Figures

FIGURE 1.1: Female shares in selected occupations in the U.S.: 1970 to 2016



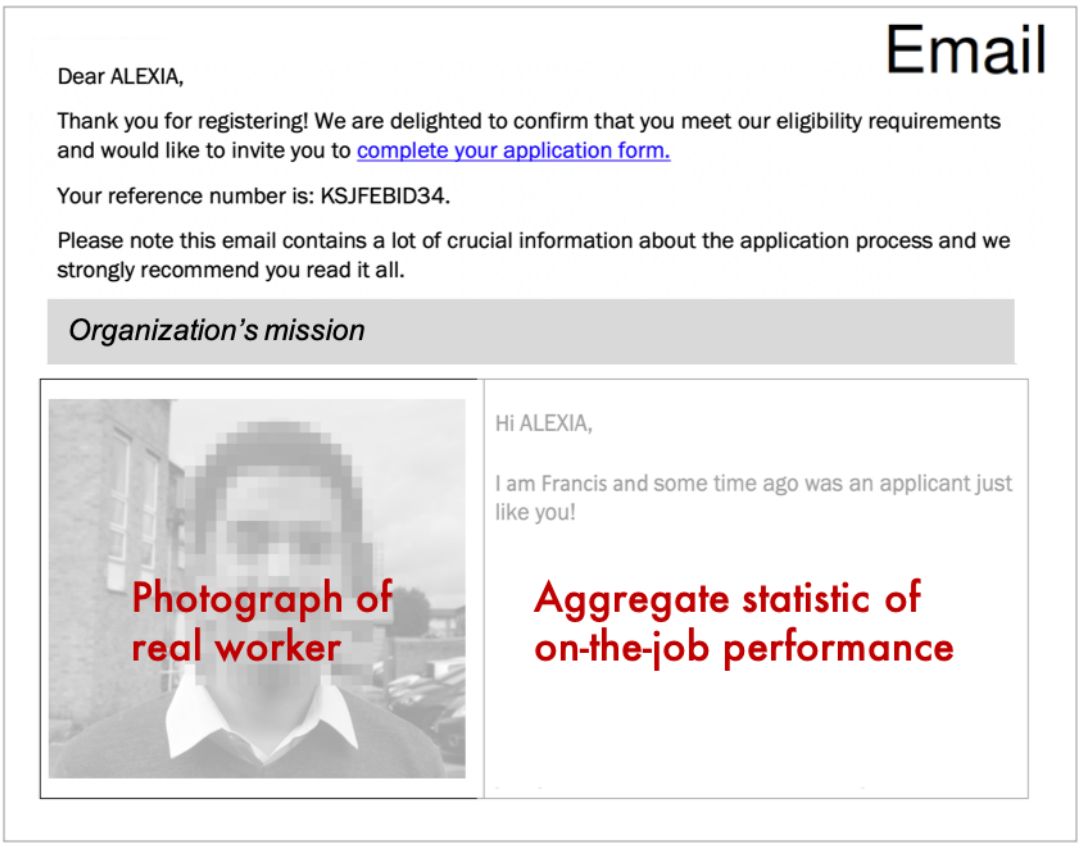
Note. Data for 1970 to 1990 are from Blau et al. (1998), who use the U.S. Census Data (U.S. Department of Commerce, Bureau of the Census) Data for 2016 are from the American Community Survey. The category "Nurses" include: licensed and registered nurses, licensed practical and vocational nurses and nursing 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 1.2: Recruitment timeline



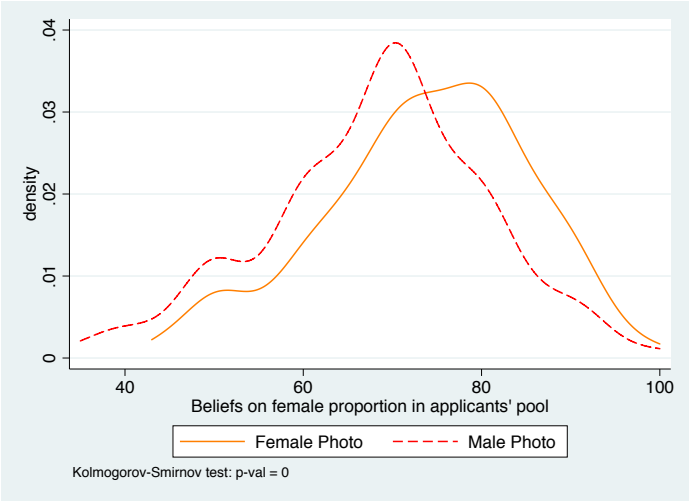
Note. The Figure shows the recruitment timeline of the partner organization from the candidates' perspective. Applications were open from September until November 2017. Randomization of the invitation-to-apply was happening between online registration and application submission. After submitting the 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 1.3: Intervention email template



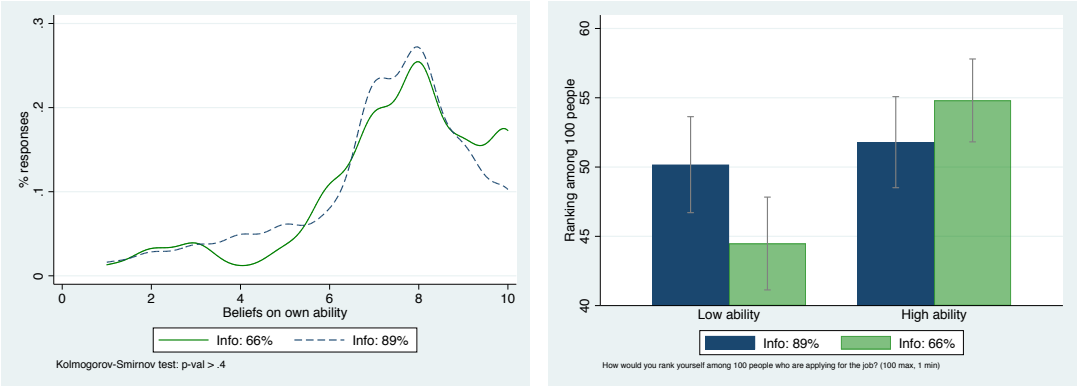
Note. The Figure shows a stylized example of one of the email templates used in the intervention.

FIGURE 1.4: Gender shares shock: manipulation checks



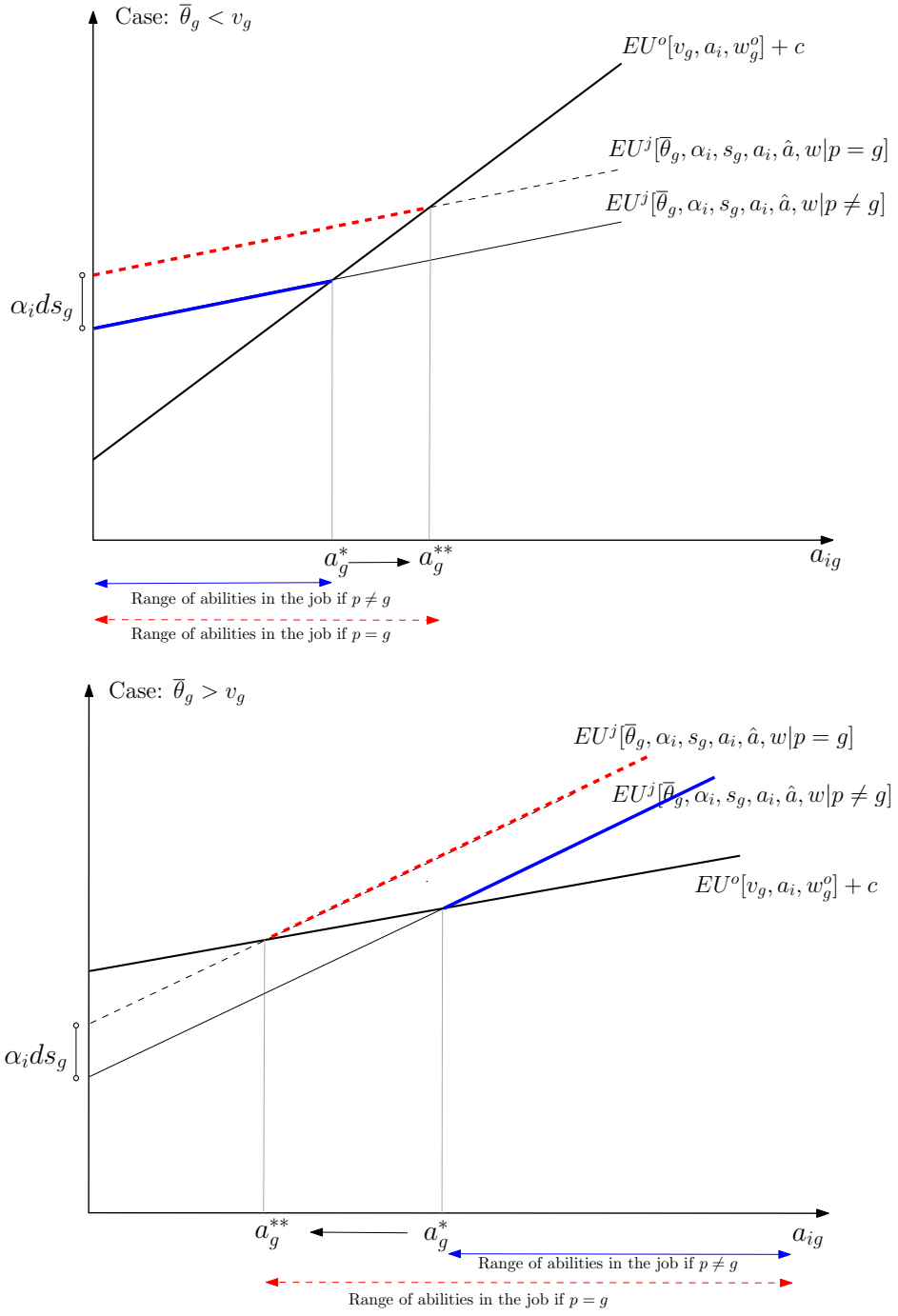
Note. The figure shows the distribution of answers to the question "Consider 100 people who apply for this job. How many do you think are women?", separately for respondents assigned to the email with a female or male photograph. Data are from the auxiliary online surveys. The dashed (solid) line is 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 1.5: Expectations shock: manipulation checks



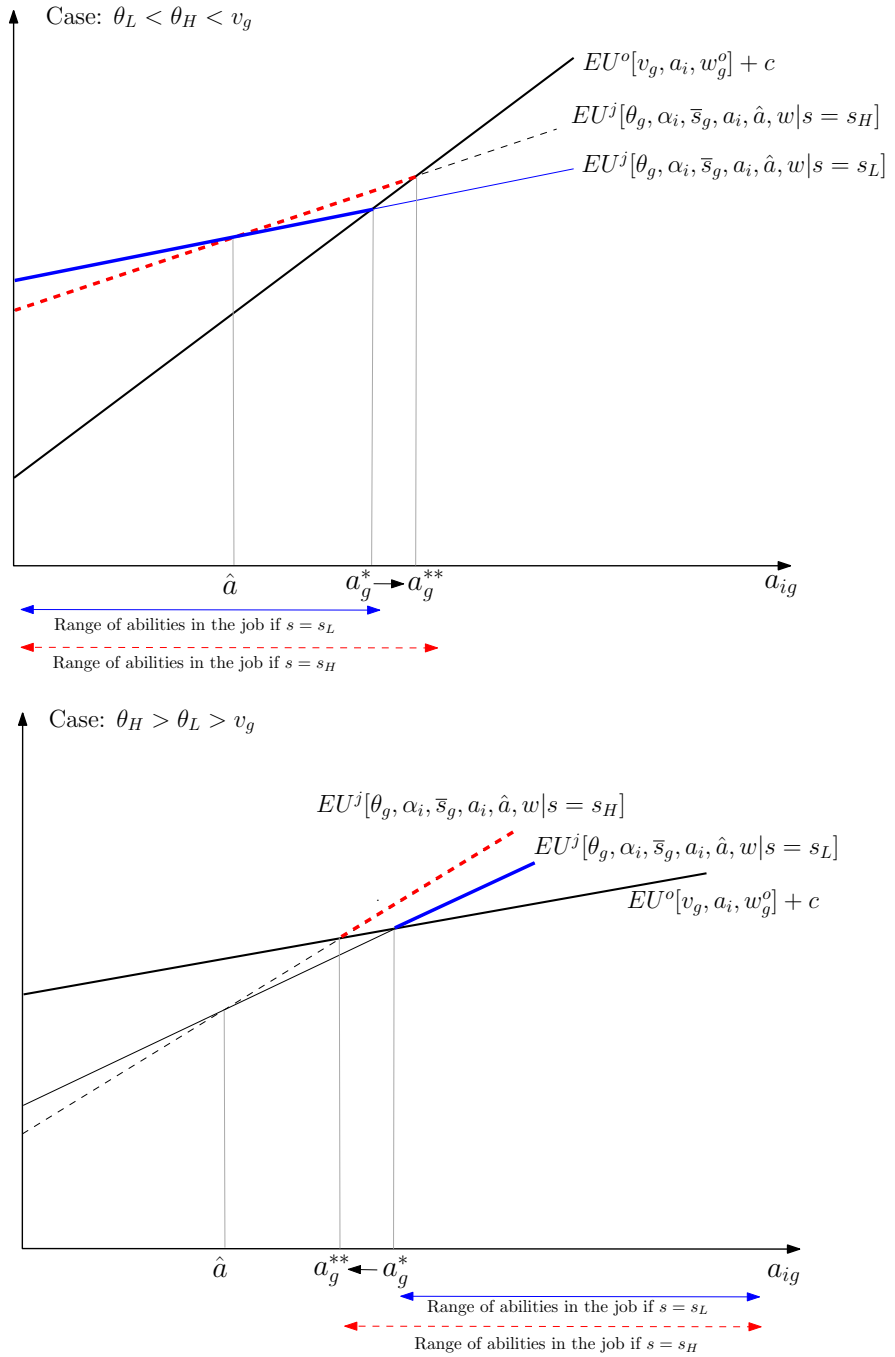
Note. The left panel shows the distribution of answers to the question "How do you expect a person with your skills and experience to perform in interacting with families in need?" on a scale from 1 (min) to 10 (max), separately for respondents assigned to the email with a statistic of 66% (solid line) 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 1.6: 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 expected utility in the outside option. The dashed and thin solid lines show the expected job utility when receiving a photo of the same $(p = g)$ or different gender $(p \neq g)$, respectively. The vertical distance between these two lines comes from the assumption of the model $E[s_g|p = g] > E[s_g|p \neq g]$. The two thresholds of ability for the marginal applicants a_g^* and a_g^{**} are determined from the intersection of the expected job utility and expected outside option. From Result 1, the size of the applicants' pool is greater when $p = g$ than $p \neq g$. 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.

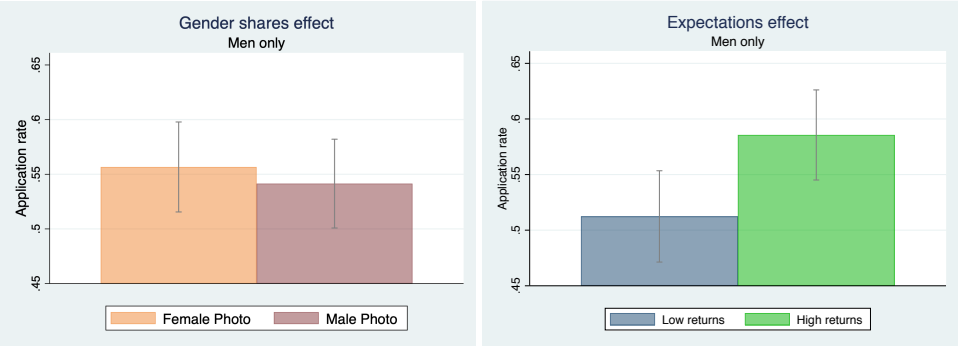
FIGURE 1.7: Theory: effect of a shock to expectations of returns to ability



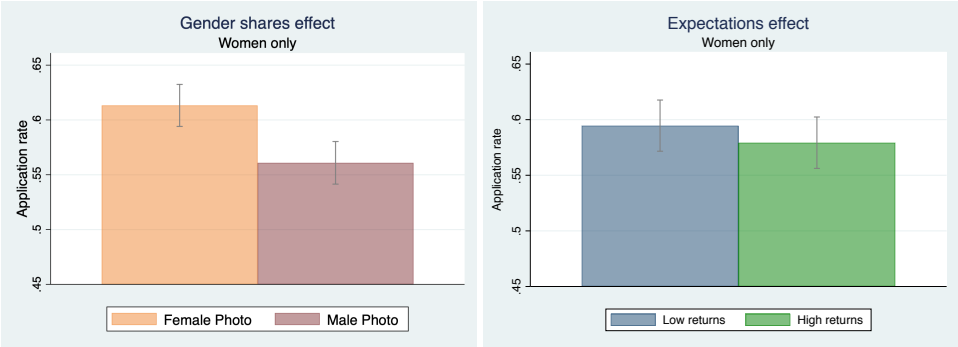
Note. The figure plots the application decision for potential applicants of gender g. The top panel considers the case $U^{j'}(a_i) < U^{o'}(a_i)$ and the bottom panel the case $U^{j'}(a_i) > U^{o'}(a_i)$. The solid black line shows expected utility in the outside option. The two thin dashed and solid 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 these two lines is explained by $E[\theta|s = s_H] > E[\theta|s = s_L]$, as higher returns to ability correspond to a higher slope. The two thresholds of ability for the marginal applicants a_g^* and a_g^{**} are determined from the intersection of the expected job utility and expected outside option. From Result 2, the applicants' pool is larger when $s = s_H$ than $s = s_L$ as long as $B > 0$ and $U^{j'}(a_i) < U^{o'}(a_i)$. 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.

FIGURE 1.8: Application rates by treatment and gender

(A) Men



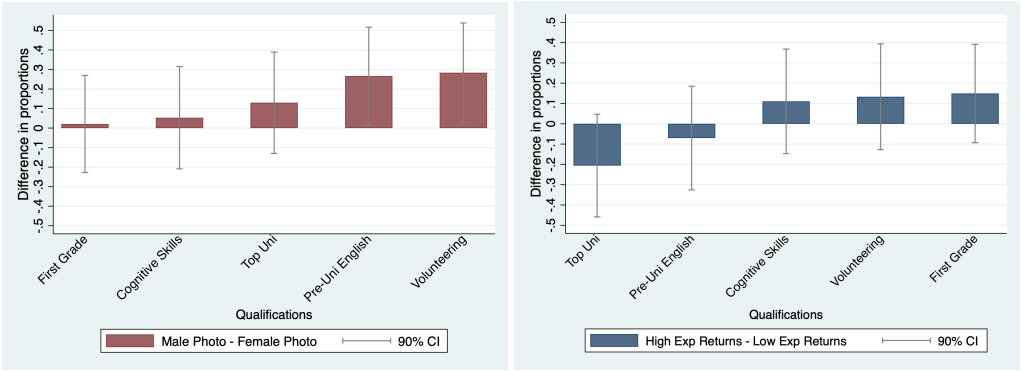
(B) Women



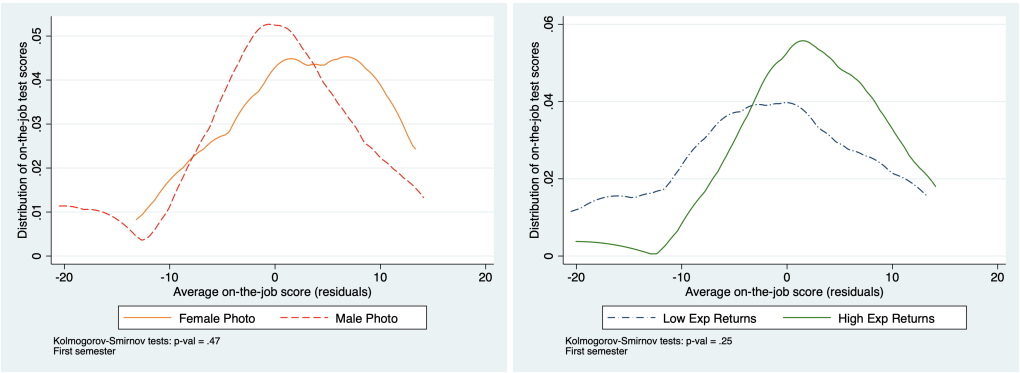
Note. Panel A shows application rates for men by photograph treatment (left-hand side) and information treatment (right-hand side). Panel B shows application rates for women by photograph treatment (left-hand side) and information treatment (right-hand side). Error bars show 95% confidence intervals.

FIGURE 1.9: Men's qualifications and average on-the-job test scores by treatment

(A) Qualifications



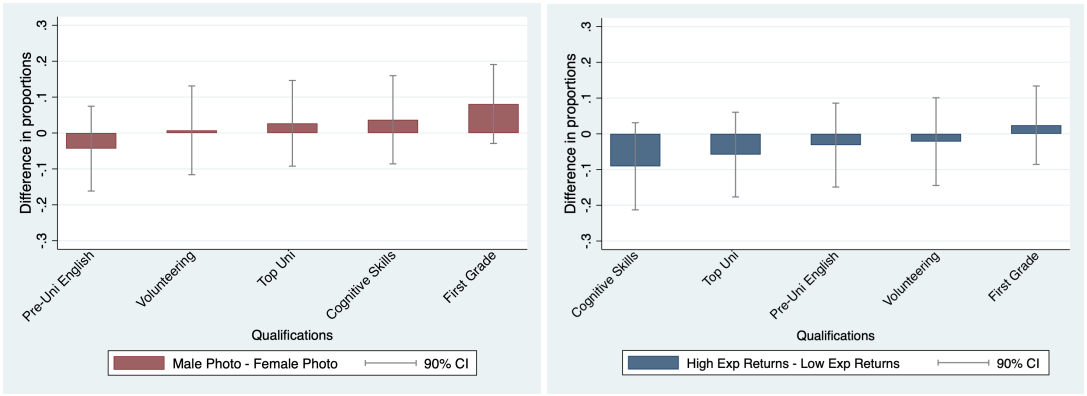
(B) Average on-the-job test scores



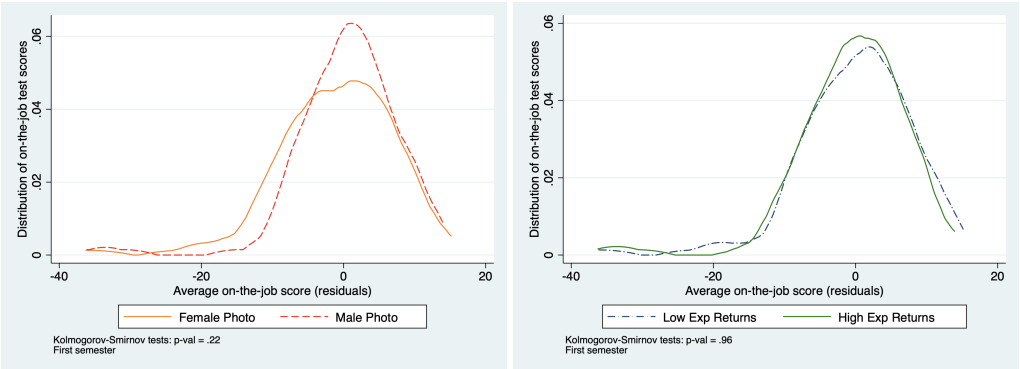
Note. The figure shows differences in the proportion of men that hold a certain qualification between treatment groups (Panel a) and the cumulative distribution of men's average test scores during the first semester on the job (Panel b), after controlling for ethnicity, past application and early registration. Figures on the left-hand side show the distributions by photograph treatment and the dashed lines are for the male photograph. Figures on the right-hand side show the distributions by information treatment and the dashed lines are for high expected returns to ability. 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 1.10: Women's qualifications and average on-the-job test scores by treatment

(A) Qualifications

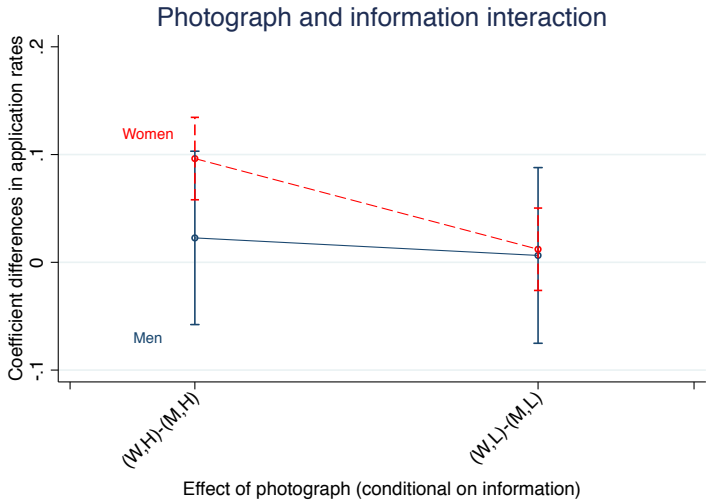


(B) Average on-the-job test scores



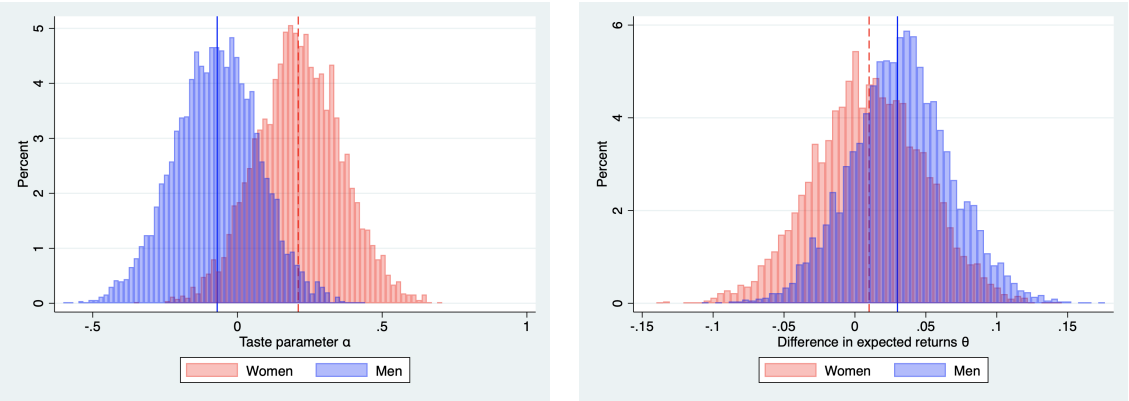
Note. The figure shows differences in the proportion of women that hold a certain qualification between treatment groups (Panel a) and the cumulative distribution of women's average test scores during the first semester on the job (Panel b), after controlling for ethnicity, past application and early registration. Figures on the left-hand side show the distributions by photograph treatment and the dashed lines are for the male photograph. Figures on the right-hand side show the distributions by information treatment and the dashed lines are for high expected returns to ability. 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 1.11: 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 1.12: 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.

1.15 Tables

TABLE 1.1: 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 1.2: 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	0.52	0.10	0.70	0.13
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 (1.2) of Section 1.5.1). 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 1.3: 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	0.48	0.57	0.53	0.52
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 (1.2) of Section 1.5.1). 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 1.4: 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	0.45	0.55	0.49	0.52
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 (1.2) of Section 1.5.1). 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 1.5: Employer's hiring criteria

	Information		Photographs	
DV:	(1) Offer	p-val	(2) Offer	p-val
Top University * T^1	0.054*		0.070**	
	(0.028)	0.67	(0.028)	0.69
Top University * T^2	0.071**		0.054*	
	(0.029)		(0.029)	
First Grade * T^1	0.110***		0.063**	
	(0.032)	0.19	(0.030)	0.04
First Grade * T^2	0.109***		0.160***	
	(0.032)		(0.034)	
Aligned Subject * T^1	-0.010		0.007	
	(0.019)	0.06	(0.019)	0.75
Aligned Subject * T^2	0.029		0.013	
	(0.020)		(0.020)	
Past Volunteering * T^1	0.047**		0.053***	
	(0.020)	0.76	(0.020)	0.85
Past Volunteering * T^2	0.056***		0.048**	
	(0.020)		(0.020)	
Maths Pre-Uni Score * T^1	0.004		-0.029	
	(0.027)	0.31	(0.024)	0.38
Maths Pre-Uni Score * T^2	-0.033		0.004	
	(0.026)		(0.029)	
English Pre-Uni Score * T^1	0.084***		0.089***	
	(0.025)	0.57	(0.024)	0.34
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 1.6: 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 (1.2) of Section 1.5.1). 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 1.7: Perceived social impact and intent to stay

DV:	(1) Perceived impact at work	(2) impact outside	(3) Confidence own practice	(4) Recommend programme	(5) Intent to stay in LA	(6) to stay 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 Dep Var	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 (1.2) of Section 1.5.1). "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 1.8: 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	0.60	0.14	0.55	-0.21
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 (1.2) of Section 1.5.1). 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 1.9: 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 1.10: Replicating the experiment in a male-dominated sector: results from a pilot

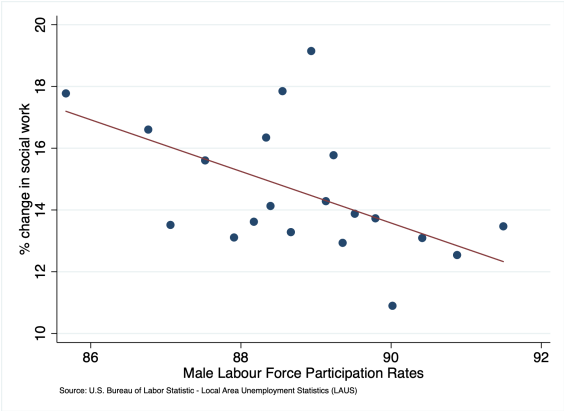
DV: Applied = 1			
	(1)	(2)	(3)
		Women	Men
High Exp Returns	0.023*	0.034	0.015
	(0.014)	(0.023)	(0.022)
Observations	900	408	492
R-squared	0.039	0.058	0.055
Basic Controls	Y	Y	Y
Country FE	Y	Y	Y
Mean Dep Var	0.05	0.05	0.05

Robust 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.

1.A Appendix figures and tables

FIGURE 1.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 1.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	0.52	0.50	0.51

Robust 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 (1.2) of Section 1.5.1). "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 1.A.2: Treatment effects: photographs and information interacted

DV: Applied and never DO = 1		
	(1) Men	(2) 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 (1.3) of Section 1.5.1). 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 1.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	0.60	0.14	0.55

Robust standard errors in parentheses

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

Note. OLS estimates for the pooled sample of men and women. The omitted category is the treatment group which received the female photograph and information 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 (1.2) of Section 1.5.1). 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 1.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	0.56	0.71	0.58	0.60	0.51
Photo = Exp Ret p-val	0.10	0.86	0.50	0.56	0.29	0.09
Rand Inf p-val						
Male Photo	0.00	0.60	0.05	0.76	0.04	0.04
Exp Returns	0.34	0.77	0.31	0.29	0.663	0.73

Robust standard errors in parentheses

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

Note. OLS estimates for women only. The table reports results of five different regressions. The omitted category is the treatment group which received the female photograph and information 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 (1.2) of Section 1.5.1). 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 1.A.5: Treatment effects by regional wage dispersion

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

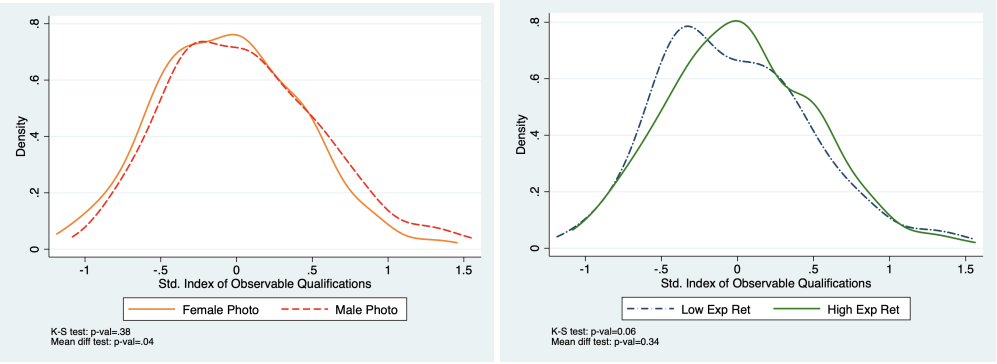
Robust standard errors in parentheses

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

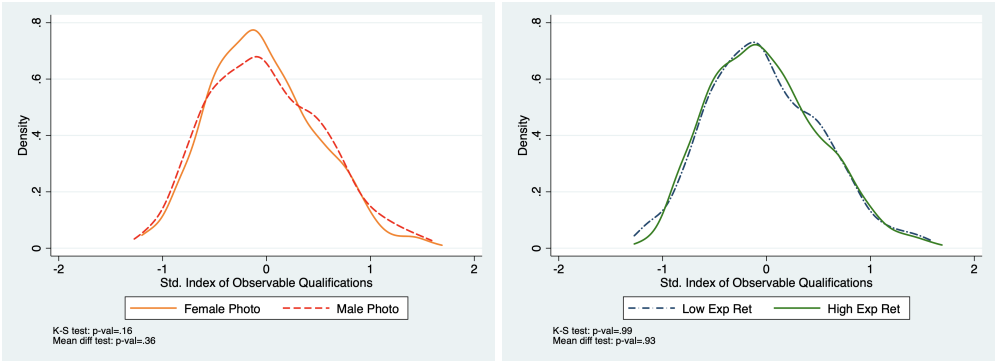
Note. OLS estimates run separately for men (Columns 1 and 2) and women (Columns 3 and 4). "Wage dispersion" is computed as the 75/25 interquartile range of the gender-specific distribution of hourly wages across industries in the UK region where the candidate lives. The level "high" or "low" is defined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

FIGURE 1.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 photograph treatment and the dashed lines are for the male photograph. Figures on the right-hand side show the distributions by information treatment and the dashed lines are for 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 1.A.6: Effort in application completion

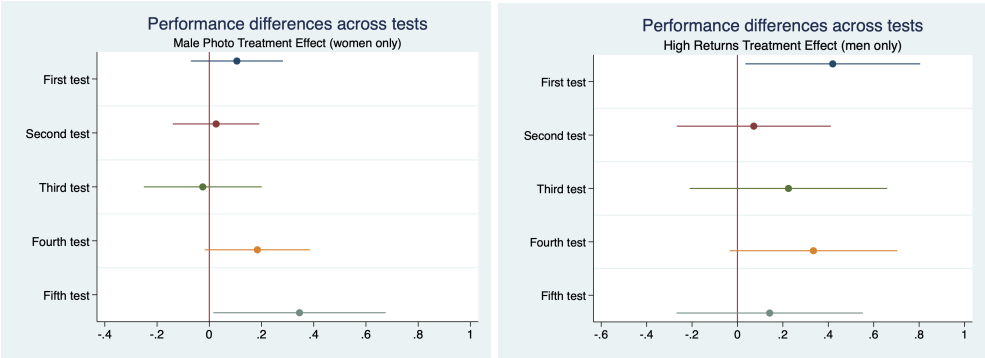
VARIABLES	(1) Access to portal	(2) # edits	(3) % completed	(4) Qst 1 length	(5) Qst 2 length
High Exp Returns	0.009	4.373**	0.026	34.178	42.509
	(0.025)	(1.970)	(0.023)	(55.655)	(46.081)
Observations	804	807	807	807	807
R-squared	0.033	0.043	0.030	0.022	0.027
Basic Controls	Y	Y	Y	Y	Y
Week dummies	Y	Y	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 (1.2) of Section 1.5.1). All the regressions control for the basic set of controls X_i: past application, access to early registration and non-white ethnicity.

FIGURE 1.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 1.A.7: Women's on-the-job performance

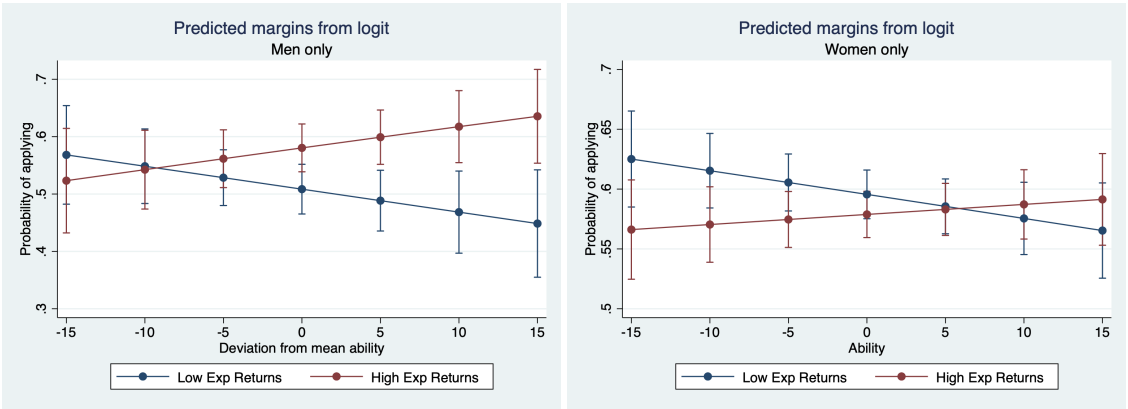
DV: First Semester Std. Scores		
	(1)	(2)
Male Photo	0.126^{+}	0.168*
	(0.082)	(0.091)
High Exp Returns	-0.031	-0.025
	(0.080)	(0.068)
Observations	955	955
R-squared	0.132	0.131
Basic Controls	Y	Y
Mean Dep Var	0.07	0.07

Clustered standard errors in parentheses (ind. level)

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

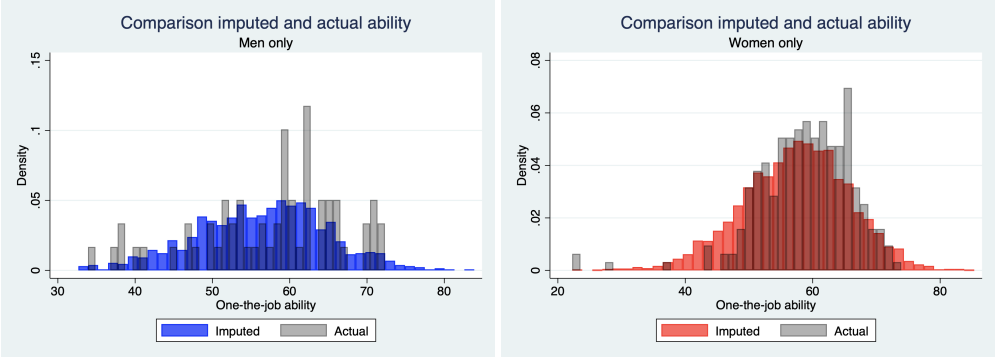
Note. OLS panel estimates for women only. The table reports results of two different regressions. The omitted category is the treatment group which received the female photograph and information of low returns to ability. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information of high returns to ability (specification (1.2) of Section 1.5.1). 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 1.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 1.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 1.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	0.84	0.84	0.80	0.80

Robust standard errors in parentheses

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

Note. OLS estimates. The dependent variables is a dummy equal to one if the candidate never asked for a reminder of his/her unique candidate number, which is needed to access the application portal and is shown in the invitation-to-apply email. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment and the regressor "High Exp Returns" is a dummy equal to one for receiving information of high expected returns to ability (specification (1.2) of Section 1.5.1). 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 1.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	0.60	0.61	0.53	0.54

Robust standard errors in parentheses

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

Note. OLS estimates. The regressor "Male Photo" is a dummy equal to one for the male photograph treatment. 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 1.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 1.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.

1.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.

1.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]

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

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

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.

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 1.B.1 presents mean differences between the male and female photographs within

[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.

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 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 1.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 1.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 1.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 1.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 1.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

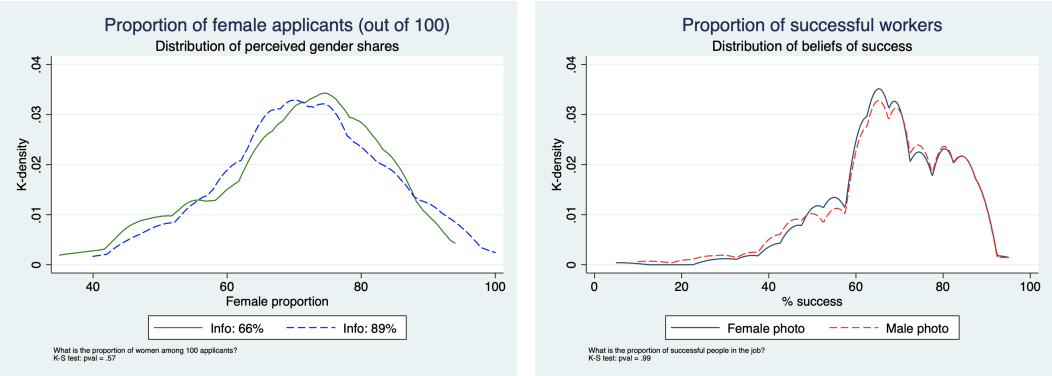
TABLE 1.B.1: Photographs: manipulation checks

	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.

treatments are interacting, which would make hard the identification of the two separate channels.

FIGURE 1.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.

1.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 1.10).

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 1.3).

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 1.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 1.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.

1.C Exposure to occupational gender segregation

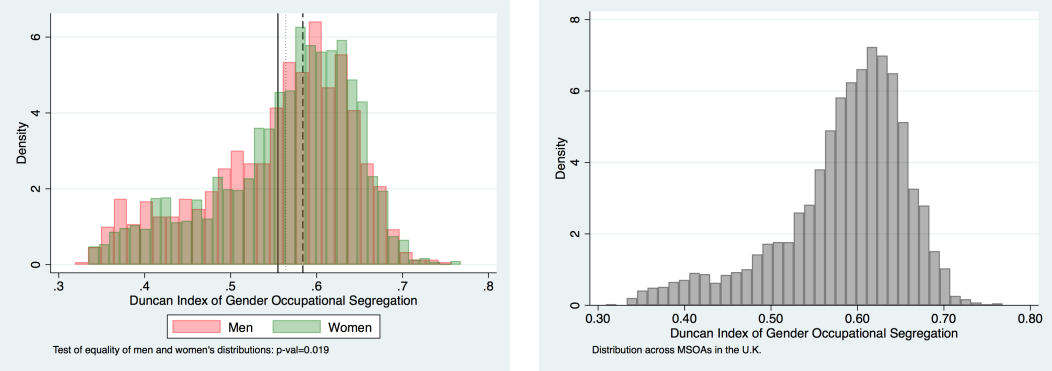
1.C.1 Measures and methods

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 1.C.1. The U.K. average Duncan Index across MSOAs is 0.5839 and the average in my sample is 0.563. Table 1.C.1 shows demographic characteristics by gender and exposure to high versus low gender segregation.

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

FIGURE 1.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).

TABLE 1.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.

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

The validity of the proxies for α used in Section 1.6.3 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 1.C.2 shows the distribution of d-score for women (left panel) and men (right

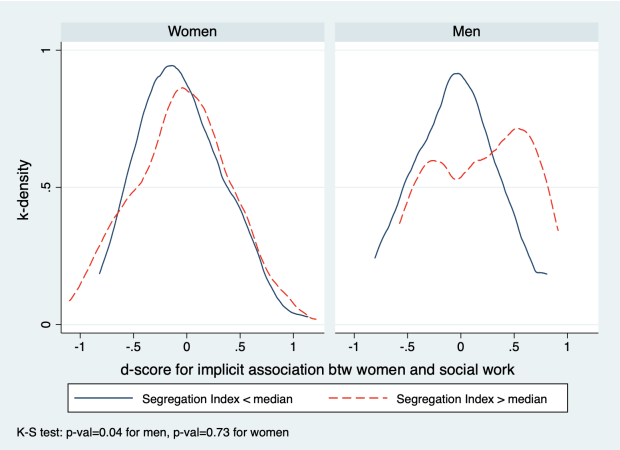
[108] Response rate was 12.5% for the main survey and 6% to the IAT (604 and 300 respondents respectively).

[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.

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

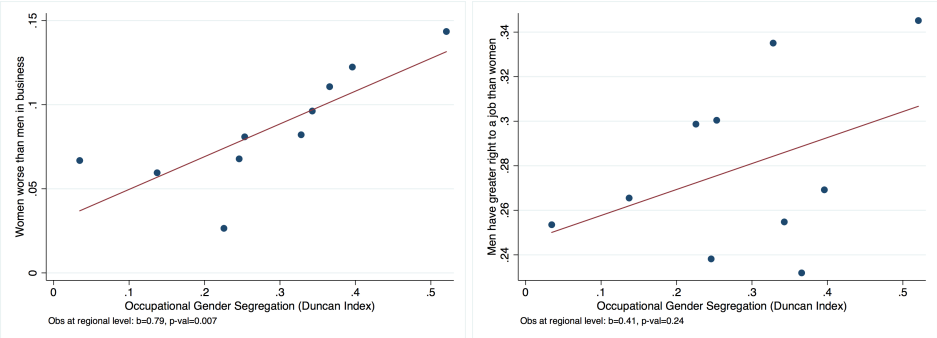
FIGURE 1.C.2: Implicit Association Test and exposure to gender occupational segregation



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 1.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 1.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.

FIGURE 1.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.

Table 1.C.2 uses data from the auxiliary online experiments (described in Section 1.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 variable in Table 1.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 1.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 Dep Var	74.66	80.18

Robust standard errors in parentheses

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

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

1.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 1.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 1.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
Scientific Subject	.15	.36	.32	.47	0	0.05	0.09
Aligned Subject	.44	.5	.27	.45	0	0.70	0.48

Note. The first five Columns of the table show summary statistics from a random sample of the LFS which I generated to reproduce the same age distribution of the experimental sample. 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.

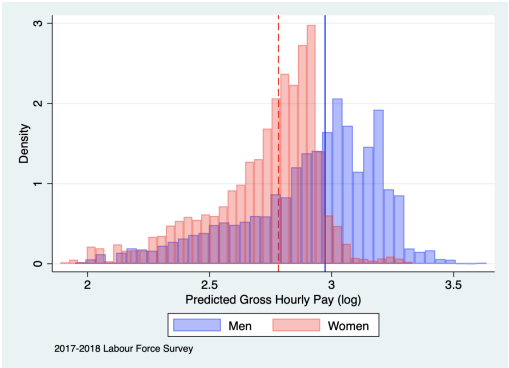
[112] For instance, the LFS data do not contain the exact university attended by the respondents.

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 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 1.D.1 shows the distribution of the computed outside option by gender. Table 1.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 1.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 1.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.

1.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 conditional quantile function $Q(score_{ia}|X_i) = \alpha + \beta T_i$, where 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 1.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 1.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	Y	Y	Y	Y	Y
Controls for Quality	Y	Y	Y	Y	Y
Clustered s.e. in parentheses (ind level)					
*** p<0.01, ** p<0.05, * p<0.1					

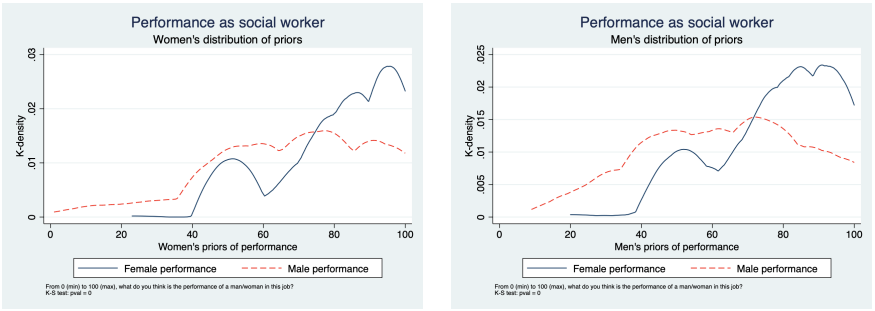
Note. Quantile regression with panel data. Estimations are for women in the top panel and for men in the bottom panel. The omitted category is the treatment group that received the female photograph and the low returns information. The regressor "Male Photo" is a dummy equal to one for the male photograph 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.

1.F Appendix to theoretical framework

1.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 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 0 is for extremely bad, 50 for neither bad nor good and 100 for extremely good performance. Figure 1.F.1 shows the distribution of beliefs held by women (left) and men (right). Both genders think that men have on average a lower performance in social work, which supports the assumption $\theta_M < \theta_W$. The variance of the distribution of beliefs about men is greater than the one of the distribution of beliefs about women, which supports the assumption $\overline{\sigma}^2_M > \overline{\sigma}^2_W$.

FIGURE 1.F.1: Beliefs about performance in social work by gender



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. Data are from the auxiliary online surveys.

Table 1.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 related to social work, the assumption of known (or unbiased expectations of) a_i seems appropriate.

TABLE 1.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.

1.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.

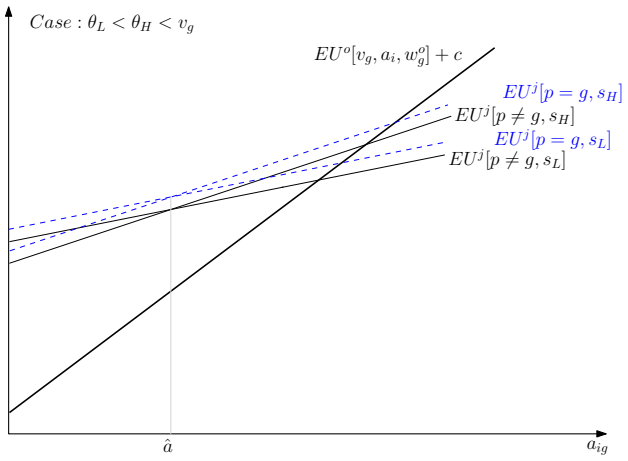
Result 3. ***Interaction between gender shares and expectations***

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

FIGURE 1.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.

1.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., 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

[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 1.10). 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.

[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 1.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.

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 ($p \neq g, s = s_H$).

1.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 $[a_1, a_2]$

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.

□

1.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 1.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.

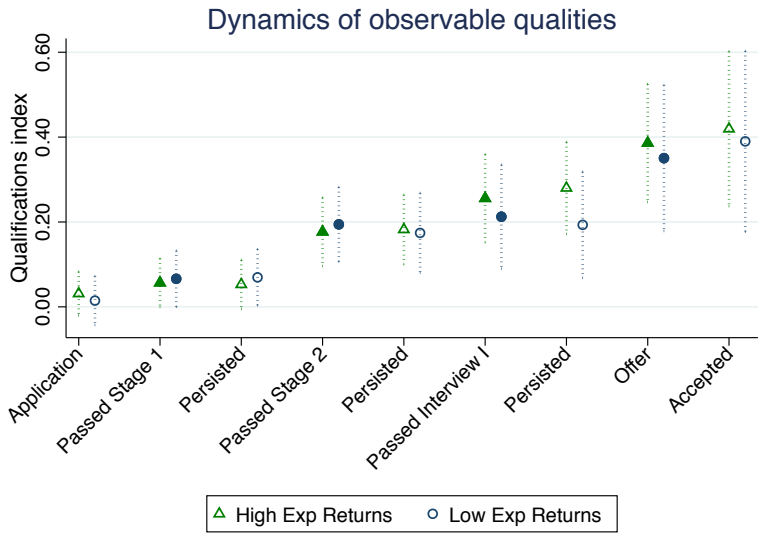
FIGURE 1.G.1: Dynamics: stayers over the hiring process (men only)

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 1.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 1.A.2 and in Table 1.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 1.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.

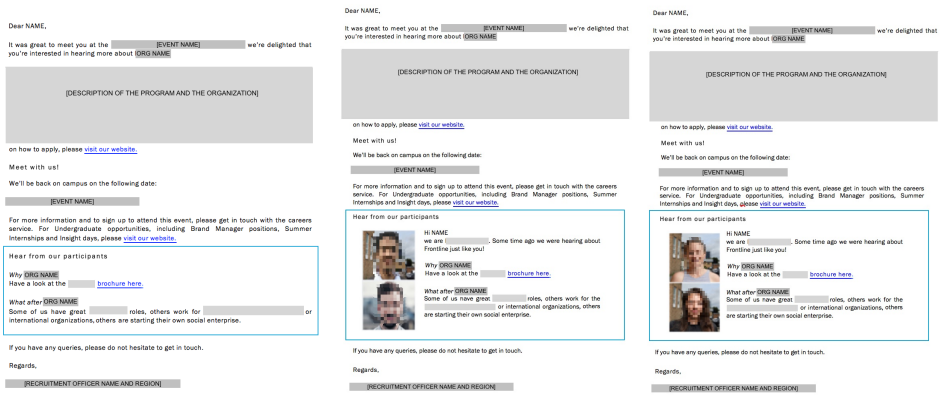
Figure 1.G.2 shows the average proportion of "desirable" observable qualifications that men have in each stage of the hiring process, by information treatment. Full symbols refer to the four stages which involve screening by the employer, as in the previous Figure. Hollow symbols refer to intermediate 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 1.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.

1.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 1.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 1.H.1. Assignment to treatment was stratified by university, event and gender.

[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.

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 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 1.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 1.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 1.H.2). Despite this first stage, behaviour does not translate into more applications. Table 1.H.2 estimates the effect of each of the treatment emails on application for people of

[119] To be trackable, unique links at the university-event-gender-treatment level were created before the randomization by adding an alpha numeric snippet to the website url.

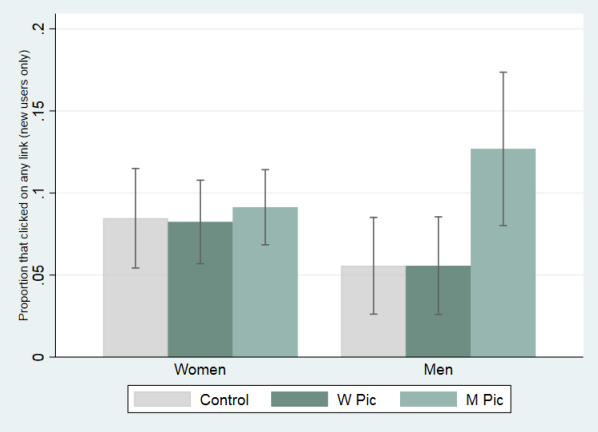
[120] Mean participation covers substantial variation between event types: stands at career fairs had an average number of attendees around 44 compared to an average of 16 for presentations and panel events.

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 1.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.

FIGURE 1.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 1.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 Dep Var	0.082	0.17

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

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

1.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 1.C.2 to construct a proxy of individual overprecision 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 1.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 overprecision 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 1.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 Dep Var	0.53	0.56	0.48	0.62

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

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

2 Rule of Law and Female Entrepreneurship

2.1 Introduction

When rule of law is weak, asymmetries of social power and physical strength stymie the trust and trade that make cities productive. The shadow of violence looms behind aggressive bargaining in many negotiations, and violence is disproportionately male (Wilson and Herrnstein, 1985). Consequently, women may avoid transacting with men unless they have the protection of courts and civil society. Around the world, women report trusting others significantly less than men do, particularly in developing countries with weaker legal institutions.[1]

In this paper, we ask whether limitations on the rule of law can explain why entrepreneurship remains a male-dominated activity in many countries and why female entrepreneurs often cluster in industries filled with other women (OECD, 2012; Fairlie et al., 2017; Singer et al., 2018; Campos et al., 2019).[2] Clustering in industries, such as apparel and food production, that contain female collaborators and customers, allows female entrepreneurs to avoid interactions with men who may have a power advantage in both legal and illegal conflict. This segregation may explain why women appear to receive lower benefits from loans (de Mel et al., 2008, 2009) and business training (de Mel et al, 2014; Bergeet al., 2015).[3]

In Section II of this paper, we present a model that extends Behrer, Glaeser, Ponzetto and Shleifer's (2019) result that when courts can be subverted, the strong trade only with the strong and the weak trade only with the weak. In our model, women choose whether to become entrepreneurs and whether to partner with men. The model embeds two aspects of inter-gender trading relationships. When courts are weak and facts are unclear, then adjudicators favour the powerful and men are more likely to have power. Even

[1] A large literature documents a strong association between trust and economic growth (Knack and Keefer 1997; Guiso et al., 2006; Algan and Cahuc, 2010). Trust is lower in developing countries, where rule of law is weak. Trust is particularly low among women in weak rule-of-law countries, including Zambia. For example, in Africa and South America, typically between ten and fifteen percent of responds say that most people can be trusted, but that number falls on average by 6 percent among women (WVS). Trust has also been shown to be especially low among the poor (Bachas et al., 2017).

[2] Worldwide the proportion of female ownership is below 50 percent and in half of the countries the female proportion is below 20 percent (see Section III).

[3] At the same time, it seems consistent with recent evidence that women might benefit more than men on training in non-cognitive skills (Attanasio et al., 2011; Acevedo et al., 2018).

when courts are strong, if social norms favour male aggression, then men can extract a greater share of the rents from bilateral commercial transactions.

For some parameter values, men actually prefer weak institutions that reduce overall social surplus, because those institutions increase the male share of profits. For other parameter values, male bargaining power and legal strength becomes a liability for men as well as women, because women refuse to partner with men. Rule of law increases the returns to female entrepreneurship, because anarchy privileges male aggression and violence. The model predicts that female entrepreneurs will earn less and segregate into predominantly female industries, unless institutions are strong and social norms favour gender equity.

We test the implications of this model both with cross-national data and by focusing on female entrepreneurs in Zambia. In Section III, we used the World Bank Enterprise Survey, the World Justice Project, and the World Values Survey to test the predictions of our model. We document that female entrepreneurship is rare in most countries, and especially in those places where rule of law is weak. The World Justice Project data indicates that courts particularly discriminate against women in many countries and that rule of law is more strongly correlated with female entrepreneurship when courts are deemed, by the World Justice Project, to be less discriminatory towards women.

We measure gender norms against women with long-standing discriminatory practices within families and limited protection of female physical integrity. Female entrepreneurship is rarer when these measures are high. As the model predicts, female entrepreneurship is more common in societies that have both rule of law and gender norms that favour women.

We also document that female entrepreneurs segregate into industries, such as hospitality, food and apparel, where they cooperate primarily with other women. The self-selection of female entrepreneurs into less profitable activities is pervasive in developing world cities (Klapper and Parker, 2011; Campos et al., 2019). Again, as the model predicts, we find that there is more female entry into male dominated industries in countries with better rule of law, less discrimination against women and especially in countries that have both.

In Section IV, we turn to our Census of Entrepreneurs in Lusaka, Zambia. Zambia is a country with both weak rule of law and discriminatory gender norms.[4]

We collect geocoded data on more than 2000 firms, which represents sixty percent of all the manufacturers in Lusaka. Interviews suggest that economies of scale can generate large returns to collaboration for these entrepreneurs. In our sample, twenty-seven percent of the entrepreneurs in manufacturing are women, and women earn slightly more than one-half of male earnings. In Lusaka, three-fourths of female entrepreneurs make

[4]According to the World Justice Project (link), Zambian rule of law is neither particularly good nor particularly bad for sub-Saharan Africa. The country's score on the World Justice Project's Rule of Law Index in below Ghana and South Africa, but above Zimbabwe and Nigeria, and is about the same as Russia and Mexico.

apparel and eighteen percent make food. Between one-half and three-fourths of the gender earnings gap for Lusaka entrepreneurs can be explained, in an accounting sense, by the massing of female entrepreneurs into two industries that are neither capital nor trust intensive. While many factors contribute to the segregation of women in these industries, our qualitative work shows that Lusaka's female entrepreneurs themselves emphasized the difficulties of trusting men.[5]

In Section V, we present our survey measures of trusting behaviour, such as working collaboratively to fill an order or jointly buying inputs or even giving advice. As the model predicts, women are less likely to take actions that require trust. Perhaps most strikingly, we find that women are much less likely to learn their trade from incumbent workers. Instead, they turn more often to formal educational institutions. Alfred Marshall (1890) emphasized that in dense clusters, "the mysteries of the trade become no mystery but are, as it were, in the air," but it seems as if female entrepreneurs cannot access these human capital spillovers in Lusaka, partially because they cannot trust the men that surround them.

We then test whether Lusaka's female entrepreneurs trust more when rule of law is stronger. We focus on the two major local institutions that adjudicate commercial disputes: Market Chiefs and Small Claims Courts (SCC). The Small Claims Court is a new institution that enables individuals with small lawsuits to bypass Zambia's overloaded and cumbersome court system. We measure institutional strength by proximity to the Small Claims Court and location within a market that is adjudicated by a chief.

Female-led businesses located inside a formal market or closer to the Small Claims Court (SCC) cooperate more, even controlling for business density, industry and other area and business characteristics. Locating within a market is also correlated with higher sales for female-led businesses. While locating within a market is an endogenous decision, the Small Claims Court was established more recently and proximity to a court is less salient than membership in a market. Proximity to the Small Claims Court results may therefore provide a more plausible estimate of the causal impact of institutions on the gender gap in entrepreneurial cooperation.

To address this endogeneity issue in Section 5, we present our population of entrepreneurs with an adapted version of the trust game (Berg et al, 1995), framed as an opportunity to invest in another person's business. We randomize pairs of players into three groups: a control group that received no access to institutions, a treatment group that had access to the Small Claims Court, and a second treatment group that had access to the market chief. As the overwhelming majority of our respondent had either not heard of the small claims court or thought that it was not useful for people like them, we focus on the results with the market chiefs.

[5]Zambia's female entrepreneurs both say that they trust others less, and indeed are less trusting in standard laboratory measures.

In the control group, the game replicates our survey results and echoes the model. We find a significant gender gap in both trust and trustworthiness: women send fewer tokens and return fewer token than men in our game. Sending money in the trust game is significantly positively correlated with our actual measures of cooperation by the participants in their real lives.

In the game, we test whether rule of law has a disproportionate impact on women by introducing a form of adjudication modelled on actual Zambian institutions. We allow some subjects to access a market chief, ensuring an anonymous, and therefore unbiased judgement. Randomizing access to this unbiased known institution significantly increases women's trusting behaviour, increasing the surplus for both parties and earnings for both women and men.

Section 2 presents our model. Section 3 uses international data to test the implications of our model and particularly the complementarity between rule of law and female bargaining power. Section 4 describes our Zambian empirical setting and data. Section 5 presents correlational evidence on the relationship between institutions, trusting behaviour and the gender gap in entrepreneurship among small- scale entrepreneurs. Section 6 presents the lab-in-the- field evidence on the impact of rule of law on trust and business outcomes. Section 7 concludes.

2.2 Gender bias, contract enforcement and female entrepreneurship

We now present a model where female entrepreneurs first choose whether or not to enter into an industry, and then potentially to partner with men and produce. As in Behrer et al. (2019), weak legal institutions intrinsically favour the socially powerful who are able to pressure courts and judges when facts are unclear. The weak anticipate the courts behaviour and consequently avoid dealing with the strong. If men have a comparative advantage in coercing weak institutions, then women will not contract with men in weak institutional environments, and they may avoid male-dominated industries altogether. Stronger rule of law does enable women to better enforce contracts against men, but even when rule of law is perfect, male bargaining power may still limit the returns to female entrepreneurship.

In stage 0, a potential female entrepreneur "E" has the option to pay a fixed cost and enter industry i, where the share of male incumbents equals m_i. This fixed cost is paid at time 0 and is immaterial to subsequent bargaining.

In stage 1, E is offered a business opportunity to make and sell a product of value π. E is also matched with a randomly chosen potential partner "P". Neither E nor P can make the product on their own at a cost less than π. If the parties do not partner, the opportunity disappears and there are no further payoffs to either player. If the parties do partner, they create a contract that specifies a share of π, denoted "s" that will be given

to P if the contract is not breached. By assumption, courts will only enforce contracts in which s lies between zero and one, because the court's power is limited to splitting the profits.

In stage 2, P chooses to work or shirk. P can fulfill the contract and pay a cost of $q < .5\pi$ for effort and materials. He can breach the contract and pay only $q - b$, and we assume that $.5\pi > b - q$. This breach might take the form of P doing shoddy work that must be fixed by E, or not working at all (in which case $b = q$), or of P stealing E's inputs (in which case b may be greater than q). E also pays a cost of q and works during this period.[6]

In stage 3, E learns whether P breached or fulfilled the contract. If P fulfilled the contract then no further work is needed. If P breached the contract, then E must pay a remediation cost of $b + \Delta$. The value of b and Δ are both known at the time of the contract and $\Delta > 0$. If E remediates the harm, then she receives a payment of π, and chooses how much to pay P out of that sum. We assume that $2q + \Delta > \pi > b + \Delta$, so that E will remediate if work has begun, but that a partnership will not generate a positive social surplus if shirking always occurs.

In stage 4, either P or E can sue the other in a court. As in Behrer et al. (2019), the rule of law depends on the relative power of the litigants and the obviousness of the facts. Courts always enforce indisputable facts, but when facts are disputable courts favour the more powerful. We assume that courts pay a penalty for ignoring indisputable facts, such as external embarrassment or judicial review, but ignoring disputable facts is costless. The share of profits specified by the contract is always indisputable, but P's shirking is disputable with probability δ. We interpret the variable δ as capturing both elements of this particular transaction and the institutional quality of the courts and society. When institutions are stronger, a wider range of facts will be indisputable. We assume that P learns whether his shirking will be disputable at the beginning of stage 2, and that E learns whether the shirking is disputable at the beginning of stage 3. We do not allow renegotiation at any point after stage 1.

If both litigants are women, then they are equally powerful and the contract will be enforced fairly even if the facts are disputable. The court will force E to pay P the stipulated share, but will subtract damages of $b + \Delta$ from the payment to cover the damages if shirking has occurred. The court cannot force a payment from P to E, as we assume that the court's power is limited to reallocating the surplus. If P is male, then the contract will be enforced fairly if shirking is an indisputable fact. If shirking is disputable, then court will assign no damages and force E to pay P the contractually stipulated payment.

This legal bias is one source of inequality between men and women. The second bias occurs at the point of bargaining in stage 1. If P is female, then the two agents split the total expected surplus equally. If the partner is male, then he receives a share β of the surplus, where β is determined by social norms about gender and male violence Male

[6]It is never optimal for E to shirk as she is the residual claimant of the product.

bargaining power can be micro-founded by assuming that men and women alternate making offers, and when a male offer is rebuffed, the man may harm the women in some way.

When P is female, then a partnership occurs, there is no shirking and both agents receive $.5\pi - q$. Since courts will enforce contracts fairly when both parties have equal power even when facts are disputable, P knows that she will receive no payment if she shirks. Not shirking is incentive compatible as long as $s\pi > b$. If this constraint is satisfied, then the total surplus is $\pi - 2q$. The assumption of equal bargaining power implies that both partners receive one half of this amount, which implies that $s = .5$, and since $.5\pi > b - q$, the incentive compatibility constraint holds.

When P is male, then he will always shirk when there is an opportunity for disputable shirking. Men will not shirk when shirking is indisputable as long as $s\pi > b$, and that generates an incentive compatibility constraint. Proposition 1 describes the returns to partnering with men (all propositions are proven in the Appendix):

Proposition 1. *Proposition 1: If* $\frac{(\pi-2q)}{\Delta} < \frac{q}{b} - 1$, *then there is no contract if* $\frac{(\pi-2q)}{\Delta} < \delta$. *If* $\frac{(\pi-2q)}{\Delta} > \delta$, *the contract specifies* $s = \beta - \frac{(2\beta-1)q+\beta\delta\Delta+\delta b}{\pi}$ *and provides expected welfare of* $\beta(\pi - 2q - \delta\Delta)$ *to P and* $(1-\beta)(\pi - 2q - \delta\Delta)$ *to E. If* $\frac{(\pi-2q)}{\Delta} > \frac{q}{b} - 1$, *then there is no contract if* $\frac{(\pi-q-b)}{\Delta+b} < \delta$. *If* $\frac{(\pi-q-b)}{\Delta+b} > \delta > \frac{\beta(\pi-2q)+q-b)}{\beta\Delta+b}$, *then the contract specifies* $s = \frac{\beta}{\pi}$, *providing expected welfare of* $(1+\delta)b - q$ *to P and* $\pi - q - (1+\delta)b - \delta\Delta$ *to E. If* $\delta < \frac{\beta(\pi-2q)+q-b)}{\beta\Delta+b}$, *the contract specifies* $s = \beta - \frac{(2\beta-1)q+\beta\delta\Delta+\delta b}{\pi}$, *providing expected welfare of* $\beta(\pi - 2q - \delta\Delta)$ *to P and*$(1-\beta)(\pi - 2q - \delta\Delta)$ *to E.*

The proposition describes two cases that depend on whether $\frac{(\pi-2q)}{\Delta}$ is greater or less than $\frac{q}{b} - 1$. In both cases, when δ is sufficiently high, either because courts are sufficiently weak or because shirking is intrinsically disputable, then no contract will occur. In both cases, when δ is sufficiently low, then a contract splits the surplus based on the bargaining power of men. In this region, female returns to the partnership are rising with overall profits (π), and falling with costs of production (q), costs of remediation (Δ), male bargaining power (β) and the weakness of the court system (δ). Female entrepreneurship may need both a relatively egalitarian society and strong legal institutions to thrive in male dominated industries.

When $\frac{(\pi-2q)}{\Delta} > \frac{q}{b} - 1$, and either the returns to partnership or the benefits of cheating are high, then there is also a third possibility. The female entrepreneur may effectively pay her male partner an efficiency wage to stop him from shirking when facts are indisputable. This possibility occurs for intermediate values of δ, and in that case, the returns to female partnership with a male equal $\pi - q - (1+\delta)b - \delta\Delta$, which is rising with overall profits (π), and falling with costs of production (q), costs of remediation (Δ), and the weakness of the court system (δ).

In this region, male returns are increasing with the weakness of the courts (δ) even though that weakness decreases the overall surplus from the partnership. This perverse

comparative static suggests that dominant groups, like men, may actually sometimes prefer weak institutions because that weakness ensures that they will receive a larger share of the surplus, despite the fact that institutional weakness diminishes the overall surplus.

The male returns drop discontinuously from $(1+\delta)b - q$ to zero at the point where δ exceeds $\frac{(\pi-q-b)}{\Delta+b}$. At that point, female returns hit zero, and women decide not to partner with men at all. Male strength becomes a disadvantage because women shun them. This result may help explain why African-American men struggle more than women in the labour market, if a legacy of prejudice means that customers and co-workers are more afraid of African-American men than women.

Male strength may also hurt men if women do not enter the industry at all, and we turn to that margin now. We assume that the fixed cost of entry equals θ times $.5\pi - q$, the highest returns from entrepreneurship, where $\theta < 1$. This value of $1-\theta$ captures the effective "profit" margin associated with entrepreneurship.

Proposition 2. *If* $\delta > Max[\frac{\pi-2q}{\Delta}, \frac{\pi-q-b}{\Delta+b}]$ *then women do not partner with men, and enter if and only if* $1-\theta > m_i$. *If* $\delta < Min[\frac{\pi-2q}{\Delta}, \frac{\beta(\pi-2q)+q-b}{\beta\Delta+b}]$, *then women always enter if* $1 - \frac{\theta}{2(1-\beta)} > \frac{\delta\Delta}{\pi-2q}$, *but if* $1 - \frac{\theta}{2(1-\beta)} < \frac{\delta\Delta}{\pi-2q}$, *women enter if and only if* $m_i < m^* = \frac{(\pi-2q)(1-\theta)}{(\pi-2q)(2\beta-1)+2(1-\beta)\delta\Delta}$, *where* m^* *is rising with* π, *falling with* $q, \theta, \delta, \Delta$ *and* β.

Proposition 2 highlights that male bargaining power or weak institutions can both lead women to avoid any male-dominated fields. If $\delta > Max[\frac{\pi-2q}{\Delta}, \frac{\pi-q-b}{\Delta+b}]$, then if women enter they will not trade with men. Consequently, their effective profit margin $1-\theta$ must be high enough to cover the probability that they may earn nothing. If $1-\theta$ is small and institutions are weak, then women will only enter fields that are almost exclusively female.

When institutions are strong enough so that women will trade with men, then female entry can still be limited by male bargaining power. If $\delta < Min[\frac{\pi-2q}{\Delta}, \frac{\beta(\pi-2q)+q-b}{\beta\Delta+b}]$, then women and men split the surplus. Even in this case, if β is high, women will not enter fields with significant numbers of men, even if rule of law is perfect. If there is either a lack of female bargaining power or weak rule of law, then women will only enter fields where potential female partners are abundant.

Figure 2.1 illustrates the proposition's implication that female entrepreneurship can be limited by either weak rule or law or by biased gender norms. The two lines both capture the maximum share of men in an industry that women will enter. The bottom line shows the case where rule of law is weak and δ is high enough so that women will not partner with men. In that case, the maximum share is $1-\theta$ which we assume to be .1. The top line is decreasing with β and it shows the case where δ is low enough so that women will partner with men, and $\frac{\delta\Delta}{\pi-2q}$ is assumed to be .2. The threshold for entry is reasonably high when δ and β is low, but if either δ or β are high, then women will not enter into male-dominated fields.

The Appendix also details the entry condition when $\frac{\pi-2q}{\Delta} > \frac{q}{b} - 1$ and $\frac{\pi-q-b}{\Delta+b} > \delta > \frac{\beta(\pi-2q)+q-b}{\beta\Delta+b}$, so that women must effectively pay men efficiency wages to stop them from breaching the contract even when breach is indisputable. The results are quite similar, except that the threshold for female entry falls with the benefit from breach (b) since that determines the efficiency wage, and is independent of male bargaining power.

In the empirical work that follows, we test whether the existence and success of female entrepreneurship depends on rule of law, less biased social norms or both.

2.3 Cross-national evidence of female entrepreneurship and rule of law

In this section, we first document three stylized facts about female entrepreneurs globally: the rate of female entrepreneurship is often shockingly low, female entrepreneurs appear to earn less than male entrepreneurs and are strongly segregated into a small number of industries. Second, we introduce our measures of legal weakness and social bias against women. Third, we test our model's prediction that female entrepreneurship, and female entrepreneurship in male dominated fields, depends on both rule of law and limited social bias, and particularly on the co-existence of both.

2.3.1 Three stylized facts about female entrepreneurship worldwide

The relative paucity of female entrepreneurs is a well-known fact in the developed world. In the U.K., France, Germany and U.S., more than two men select into entrepreneurship for every nascent female entrepreneur according to the Global Entrepreneurship Monitor (GEM) and the Kauffman Foundation (Singer et al., 2018; Fairlie et al., 2017).[7] Gender gaps in entrepreneurship as large are also found in the developing world, but things are more heterogeneous and depend on the sector as well as geographical region. In 2016 the female-to-male ratio in entrepreneurial activity was around 0.7 in South Africa, 0.8 in Botswana and Burkina Faso and only around 0.4 in Tunisia or Egypt according to the Global Entrepreneurship Monitor (Kelley et al., 2017). The gender gap persists over the

[7]The GEM survey is collected across countries and, for each country, at least 2000 adult entrepreneurs/business managers are surveyed. The total number of surveyed people depends on "the population and the economic diversity of each country" and is supposed to be conducted on a representative national sample of adults. It interviews nascent and established entrepreneurs in urban/rural areas and excludes people who are considered to be out of the labour market (e.g., retirees). The method by which they identify participants is dictated by the percentage coverage of the landline telephone network. Where landline coverage is greater than 85 percent of all households, then the National Teams use a landline-based survey outreach to generate a suitable list of participants to contact. For those countries where landline telephone coverage is not as wide-spread, face-to-face interview techniques and/or mobile phones are used. Specific information can be found here https://www.gemconsortium.org/wiki/1157.

life-cycle of the firm and is as large for long-established firms, if not larger (Kelley et al., 2017).[8]

We replicate these results using the 2006 to 2016 waves of the World Bank Enterprise Survey. We limit our sample to businesses which are sole proprietorship or partnerships, and define female entrepreneurship as the share of firms that have a (weak) majority of female owners.[9] Figure 2.2 shows that Romania and Moldova are the only two countries in the sample with a clear majority of firms in the sample led by women (with a sample size of only 14 and 42 firms respectively). In the more than one-half of the countries, fewer than one-in-five enterprises have a majority female ownership.[10] Not only is the global level of female entrepreneurship low, the rates of female entrepreneurship appear to differ significantly across countries.

The gender gap in entrepreneurship can be associated with either decreased returns to female entrepreneurship or higher opportunity costs of women's time, perhaps because of productivity in home production. If the gender gap reflected opportunity cost of time, then the returns to entrepreneurship should be higher for women, but that does not appear to be the case. The World Bank Enterprise Survey provides more reliable measures of revenues than profits, so we focus on the revenue differences between female and male led firms. Across the entire sample, male firms average 0.6 log points higher sales than female led firms (a reduction in the geometric mean of sales by 55 percent).

Female entrepreneurs may earn less because they specialize in industries with lower returns or because there are more female entrepreneurs in poorer countries. In Figure 2.3, we show the distribution of earnings controlling for both industry (ISIC 3.1 code) and country.[11] The overall sales gender gap is 0.3 log points controlling for industry and nation (a reduction in the geometric mean of sales by 36 percent). A Kolmogorov-Smirnov test rejects the equality of these two distributions at the 99 percent level, and as the figure shows, the distributions differ especially in their right-tail. Women seem to hit a cash ceiling to their earnings.

Perhaps the most surprising fact about female entrepreneurship is that it is so concentrated in a small number of industries. The three industries with the greatest proportion of female managers across countries are retail trade (ISIC 3.1 code 52), manufacturers of food products and beverages (ISIC code 3.1 code 15) and hotels and restaurants (ISIC

[8]The Global Entrepreneurship Monitor classifies entrepreneurs in two broad categories: nascent entrepreneurs and owner-managers. A nascent entrepreneurs is someone who is involved in setting up a business. This means that the entrepreneur has been active in the past 12 months and has not paid wages in the last three month. Owner-managers are those entrepreneurs who have been working for longer and are also divided into two subcategories: owner-manager of a new firm, and owner-manager of an established firm. The former has started paying wages at most 3.5 years ago, whereas the latter has been paying wages for more than 3.5 years.

[9]We exclude 5 percent of countries with less than 10 observations in this sample.

[10]The distribution looks very similar considering female-led businesses as firms with at least a female owner.

[11]The industry fixed effects are based on the ISIC Code 3.1. This is based on the question "In the last complete fiscal year, what were this establishment's two main products (represented by the largest proportion of annual sales)?".

code 3.1 code 55). While female owners only own thirteen percent of all manufacturing firms in our global sample, they own twenty percent of enterprises in food and apparel production. The cross-country average of the Herfindahl industrial concentration index is 0.19 for female-led businesses, which is significantly greater than the 0.11 average HH for male-led businesses (p=0).[12]

The industries chosen by women not only have a greater proportion of peers of the same gender, but they also have more female customers and employees. For instance, female-led firms employ more women in both production and non-production roles than male-led firms. On average, forty-eight percent of the fulltime workforce in female-led firms is made of women. This percentage is halved in male-led firms. Similarly, sixty percent of female-owned firms have a female top-manager, as opposed to only six percent of firms with a majority of male owners.

2.3.2 Measuring rule of law and gender norms

We now turn to our measures of gender norms and rule of law. We use two measures of biased gender norms: the Social Institutions and Gender Index (SIGI) and the Global Gender Gap Index (GGGI). The SIGI Index is created by the OECD Development Centre and covers 180 countries. The World Economic Forum is responsible for the 144 country Global Gender Gap Index.

The SIGI index constructs variables on four gender-related topics (discrimination in the family, restricted physical integrity, restricted access to productive and financial resources, and restricted civil liberties) based on qualitative and quantitative data on discriminatory social institutions. SIGI relies on legal experts, government representatives, and SIGI National focal points, but not individual-level surveys. We focus on the SIGI index of discrimination in the family, which is based on laws on child marriage, household responsibilities, inheritance, and divorce. We also look at the SIGI Physical Integrity Index, which includes laws on violence against women and reproductive autonomy, attitudes towards and prevalence of female genital mutilation (FGM) and domestic violence, missing women, and access to family planning.[13]

These measures should not directly relate to commercial contracts signed by women, but they should capture social norms towards women and the social acceptability of male violence against women, which are conceptually closest to our male bargaining power variable: β. These measures are particularly removed from the commercial and public spheres, and strongly correlated with long-standing cultural practices. We supplement these measures with the proportion of respondents in a country in the World Values Survey who agree with the statement “it is justifiable for a man to beat his wife" as an added measure of gender norms about violence.

[12]For each country, the HHI is computed as the sum over industries of squared shares of women (or men) entrepreneurs.

[13]More information on the SIGI can be found here: https://www.genderindex.org/.

The GGGI index measures the progress of countries towards gender parity across four themes: Economic Participation and Opportunity, Educational Attainment, Health and Survival, and Political Empowerment. We focus on the Educational Attainment index, which is the weighted average of four variables (all in ratios): female/male literary, female/male net primary enrolment, female/male net secondary enrolment, and female/male gross tertiary enrolment. Again, our hope is that these variables capture gender norms, but do not directly relate to the enforcement of contracts signed female entrepreneurs. The four variables come from the UNESCO Institute for Statistics.[14]

For rule of law, we use data from the World Justice Project (WJP) and the World Bank's Governance Indicators.[15] We use the World Bank's rule of law index, which captures "perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence" (Kaufmann et al., 2010). We use the rank decile for each country. This measure should capture the extent to which courts will enforce binding contracts signed by men and women, which is captured by our variable δ.

To measure legal bias against women, we use the World Justice Project's survey of qualified respondents. This survey asks legal experts throughout the world to assess the quality of the judicial system. One question asks "In your opinion, how likely are the following criteria to put a person at a disadvantage before a civil or commercial trial court?" where one of the "criteria" is being female. The World Justice Project produces this measure as a score on a zero to one scale, where one represents less bias in the judicial system. We also use the overall score of the World Justice Project's survey of qualified respondents on equal treatment of the genders by all legal institutions.[16] The correlation between this measure of bias against women and the World Bank's measure of contract enforcement is .73, suggesting that when overall rule of law is weaker, courts show more bias against women, just as the model suggests.

2.3.3 Female entrepreneurship, gender norms and rule of law

We now turn to the relationship between gender norms, rule of law and female entrepreneurship. Figure 2.4 shows the cross-country relationship between female ownership, on the vertical axis, and the SIGI family discrimination measure. The correlation

[14]The weights are calculated by dividing 0.01 by the standard deviation for each indicator. More information on the GGGI can be found here: https://www.weforum.org/reports/the-global-gender-gap-report-2017.

[15]The Worldwide Governance Indicators report on six measures of governance (voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption) for over 200 countries since 1996. The six indicators are created by summarizing data from 30 different sources, that report views of citizens, experts, and private and NGO sectors. The questions from each data source used to construct the rule of law index can be found here: http://info.worldbank.org/governance/wgi/index.aspx#doc.

[16]This overall score is an average of the scores on equality of genders in courts, hospitals, government hiring and police.

coefficient is -0.62, which is significant at the one-percent level. The fitted line suggests that as a country improves from the 90th percentile in this variable (0.81) to the 10th percentile (0.22), the predicted level of female entrepreneurship increases from 7 to 25 percent. Social attitudes towards women are at least strongly correlated with low levels of female entrepreneurship.

Female entrepreneurship is also correlated with variables that measure modern, commercial institutions. The World Bank's Worldwide Governance Indicators measure the quality of contract enforcement at the national level for most of the countries in the World Bank Enterprise Survey. This measure should capture the extent to which men and women are able to sign binding agreements that enable them to work together.

Figure 2.5 shows the correlation between the World Bank index of quality of contract enforcement and the rate of female entrepreneurship across countries. The correlation coefficient is 0.32, and the relationship is statistically significant at the 1 percent level. The fitted values suggest that as a country moves from the 10th percentile rank in the quality of contract enforcement to the 90th percentile, the share of female entrepreneurship increases from 13 to 30 percent.

Table 2.C.1 shows that the univariate relationships between female entrepreneurship and our other measures of gender norms and rule of law. The violence related SIGI Index (termed the "physical integrity index") has a -.5 correlation with female entrepreneurship and a t-statistic of 3.8 in a bivariate regression with country level income.

The World Bank Global Indicators measure has a correlation of .32 and a t-statistic of 1.94 in a bivariate regression. The World Justice Project measure of court bias against women in a commercial or civil court has a correlation coefficient of .35 with the entrepreneurship measure and a t-statistic of 1.9 in the bivariate regression.

We now look at the complementarity between female bargaining power and fair and effective political institutions that was predicted by our model. Our hypothesis is that women will only enter into economic interactions with men, and most entrepreneurship requires such interactions, when they are safe both from expropriation through male bargaining power and from expropriation through the courts.

We continue to use the SIGI index of discrimination in the household, based on laws on child marriage, household responsibilities and divorce. We also use the proportion of respondents in a country in the World Values Survey who agree with the statement "it is justifiable for a man to beat his wife".

For our measures of overall legal quality (δ), we return to the World Bank's Governance Indicators measure of the quality of the rule of law at the national level years of the World Bank Enterprise Survey. We also use the World Bank's Doing Business Report measure of the ease of contract enforcement, as an alternative measure of the accessibility of the court system for contract disputes.

The first interaction that we examine is between rule of law and discrimination against women in the family. Table 2.1 shows the basic interaction. Panel A shows that there is

plenty of variation across countries in the patterns of rule of law and family discrimination. There are 47 countries in our sample with high rule of law and low discrimination in the family, and 54 countries in our sample with low rule of law and high discrimination. There are also 58 countries which are off the diagonal, with either high rule of law and high discrimination or low rule of law and low discrimination. Panel B shows the female entrepreneurship patterns across the table. The entrepreneurship rate is 16 percent in the box with low rule of law and high discrimination, but moving to either off-diagonal square increases the female entrepreneurship rate only modestly to 18 or 24 percent. Moving to the high rule of law, low discrimination panel effectively more than doubles the female entrepreneurship rate to 36 percentage points.

Table 2.2 confirms this interaction in a regression setting. In regressions (1) and (2), we look at the SIGI family discrimination figure. In regression (3) and (4), we use the measure of violence against women. Regressions (2) and (4) add controls for per capita income.

Regression (1)-(2) and (4)-(5) all show that Rule of law has little impact on female entrepreneurship in countries where families discriminate against girls or where male violence is acceptable. Regressions (1) and (2) show that where family discrimination is low, female entrepreneurship is more common even when rule of law is low. We do not find a similar independent effect of violence against women.

Our primary focus is on the interaction effects, which are positive in all four specification, and significant in three of those specification. When rule of law is combined with low levels of family discrimination or violence against women, then the impact on female entrepreneurship is extremely large. Good legal institutions seem to increase female entrepreneurship only when they are merged with a culture that allows women to extract rents from relationships with men.

Regressions (7) and (8) examine our second interaction that looks at the link between rule of law and gender bias in the courts. In regression (7) of Table 2.2, we show the impact of gender equality before the law in the World Justice Project, the World Bank Rule of Law Figure and the interaction. Once again, the interaction is more powerful than either variable on its own. Rule of law is actually negatively correlated with female entrepreneurship, unless commercial courts are seen as being unbiased against women. Regression (8) duplicates this result controlling for GDP.

Our model focused on the entry of women into male-dominated industries, and predicted that women would be willing to join female-dominated industries even when gender norms are discriminatory or rule of law is weak. While we believe that this prediction is supported by the segregation of women into a small number of industries, even these industries are typically predominantly male. Consequently, our results on overall female entrepreneurship can be interpreted as examining whether women enter into male dominated fields.

Nonetheless, we now also ask whether these rule-of-law and gender bias variables

also predict whether women enter into the industries that are much less likely to include women worldwide. To do this, we recalculate female entrepreneurship rates only in manufacturing, whose female proportion in the World Bank dataset is 13% against 17% across industries. Regressions (3), (6) and (9) of Table 2.2 use the female entrepreneurship rates in manufacturing as dependent variable. In regression (3), we look at the SIGI family discrimination figure, the World Bank Rule of Law Figure and the interaction. In regression (6), we use the measure of violence against women, the World Bank Rule of Law Figure and the interaction. Regressions (9), we show the impact of gender equality before the law in the World Justice Project, the World Bank Rule of Law Figure and the interaction. Results are very similar between the aggregate and the manufacturing sample, a fact consistent with our main hypothesis.

2.4 The Zambian context and the Census of manufacturers

In this section, we discuss the Zambian context and the Census of Manufacturers that is the starting point for our work on female entrepreneurship in Lusaka. Zambia is a natural setting to study weak institutions, gender discrimination and female entrepreneurship. Zambia ranks 80th out of 136 countries in the SIGI index of discrimination within the family and of 85th out of 109 countries in the SIGI index of physical integrity restrictions. Sixty-eight percent World Values Survey respondents in Zambia say that it is justifiable for men to beat their wives in some circumstances, which is the highest share in sub-Saharan Africa.[17] Just as in many Sub-Saharan African countries, entrepreneurship is a particularly important activity for many households in urban Zambia. Despite weak rule of law and gender discrimination, our Census still documents a sizable number of female entrepreneurs.[18]

2.4.1 The Lusaka Census of manufacturers

Between May and September 2016, we collected the Lusaka Census of Urban Entrepreneurs ("Census" from now on), which is a spatial mapping of all the firms in Lusaka. For each establishment operating from a fixed location, across all industrial sectors, the Census includes geocoded location, industrial classification at the North-American Industry Classification System (NAICS) 4-digit level, number of employees and structural description (e.g., standalone building, inside markets).[19]

[17] South Africa, Nigeria and Zimbabwe follow with 63, 54 and 51 percent respectively.

[18] In 2012, 40 percent of adults in Zambia were starting a new business according to the Global Entrepreneurship Monitor (Xavier et al., 2012). Zambia has 1.02 million informal Micro, Small, and Medium Enterprises (MSMEs)-one for every five members of Zambia's total labour force (Shah, 2012).

[19] A business was considered to operate from a fixed location if 1) the business operated from a permanent structure with concrete foundations, 2) the business maintained either stock or machinery on site overnight, and 3) the permanent structure in question was not exclusively used for storage. Our analysis split Lusaka into sub-regions called Census Supervisory Areas (CSAs), and our census covered all the businesses in 90 percent of all CSAs in Lusaka district. For security reasons, we excluded the 8 census statistical areas (CSA)

These data describe the distribution, size and characteristics of economic activities in a fast-urbanizing environment and supplements the Central Statistical Office's 2012 Economic Census of the whole country. Our Lusaka Census includes a total of 48,163 establishments. As there were 16,063 businesses listed in Lusaka District in the 2012 Economic Census, our data suggests either 200 percent growth rate over 4 years or differences in methodology or comprehensiveness. Our data includes far more small firms than the 2012 Economic Census. Ninety percent of the firms in our Census have fewer than 5 employees, six percent have been between 5 and 10 employees, and fewer than one percent having 50 or more employees.[20]

Figure 2.6 shows the spatial distribution of businesses in the Census, which enables us to construct measures of business density at a granular level.

Table 2.3 presents the distribution of businesses across industries at the NAICS 2-digit level and the main characteristics associated with the businesses. The largest sectors (by number of businesses) are retailing, accommodation and food industry, and other services (the vast majority being hair dressers). Retailing firms also typically have fewer employees.

We complemented the Census with a short survey of business owners with less than 20 employees belonging to manufacturing, mining, and construction, which we refer to as the "Manufacturers Survey". This survey had a total of 2,216 respondents, which accounts for 58.3 percent of the total population in these sectors. The survey includes questions on business practices, sales and history, levels of trust, collaborative behaviour with other businesses, and demographics.[21]We focus on manufacturing because it is traditionally male, offers the possibility of exploiting economies of scale through partnerships and has been a priority for the Zambian industrial and development strategy for decades.[22]

The manufacturers survey shows that Zambian manufacturing enterprises are smaller than those in both neighbouring and developed countries (Hsieh and Klenow, 2010). Women-led businesses represent twenty-six percent of the businesses in manufacturing, construction and mining (N=3,723), which is unsurprising given that manufacturing is often a male-dominated activity (Campos et al., 2014).[23] Women's firms have 0.38 fewer full time employees and 0.7 fewer part time employees than male-led firms.

in the region surrounding Chibolya compound (Harry Mwaanga Nkumbula ward). We also excluded areas of low population density in the following wards: Kabulonga, Lubwa, Lilayi, Munkolo, Mwebeshi, Kamulanga, Munali, Roma and Mpulungu.

[20]Data on the number of employees is not available for 17 percent of the firms, so these percentages are about the 40,517 respondents to this question. In the 2012 Economic Census, the percentages were respectively 71, 11 and 3 percent. Some larger businesses were unwilling to share their employee numbers with our census takers. In total 84 percent of businesses disclosed their employment figures.

[21]If the owner was not available, the interview was conducted with the main manager.

[22]Recent research in Uganda and Ethiopia (Campos et al., 2014; Alibhai et al., 2016) show that women could potentially have high returns in manufacturing, but both social norms and fixed set-up costs might be barriers to female entry. We focus on a different source of gender gaps in our paper, in interaction with these traditional explanations.

[23]But all in all, gender segregation is also apparent in highly developed economies, as shown in the introduction. According to the Kauffman Index of Start-up Activity 59.4 percent of new entrepreneurs in the U.S. were male in 2015 (Morelix et al., 2016).

Women earn less than men. On average, women's sales value in good weeks is 2,356 Kwacha (KW), as compared to 4,085KW for men, or 180 and 311 dollars in then current exchange rates for women and men respectively (1 KW = 0.076 USD). In bad weeks, women earn on average 599.9KW as compared to 1,313KW for men, or 45 and 100 dollars respectively. Figure 2.7 shows the kernel densities of logged-sales in good and bad weeks by gender.[24] The distributions are significantly different between men and women and women earn less most of the times (p=0.00, Kolmogorov–Smirnov equality-of-distributions test). Moreover, men have a higher variance in sales in good weeks than women's (p=0.00, variance ratio test), which is driven primarily by a thicker right-tail. While the variance in sales during bad weeks is not different between genders (p=0.40, variance ratio test), both the minimum and maximum value of sales are lower for women than for men.

2.4.2 Gender, segregation and the earnings gap

Figure 2.8 shows that Lusakan women make different sectoral choices than men. Ninety-three percent of women operate in apparel and food manufacturing, while women represent a minority in wood, metal manufacturing and printing. Women appear to select into non-complex industries, despite having on average the same qualifications as men.[25]

In our survey of manufacturers, industry choice, not observable human capital, explains much of the gender gap in earnings. Male and female entrepreneurs have similar levels of education, as shown in table 2.4. Women are more likely to have participated in management or entrepreneurship training than men, as we discuss later, and the two groups do not differ in terms of record keeping. The first two regressions in Table 2.5 show the raw gender gap and Columns (3) and (4) show that controlling for education does not reduce the gender gap in sales. These results are unchanged when using or including alternative proxies for skills, such as literacy, numeracy, social skills.[26]

Regressions (4) and (5) of Table 2.5 add controls for industry and regressions (6) and (7) add household constraints, including marital status and work time. Selection into

[24]During piloting, we found that recalling exact sales digits or for periods longer than a week was challenging for most of the respondents. We thus asked the sales in the previous working day, the sales in a good week and the sales in a bad week. If the person could not provide an exact number, we also asked for an upper and lower bound. We then imputed the average of the bounds to the sales variables.

[25]We define non-complex industries as industries that have less than the mean number of skilled occupations associated with the corresponding NAICS 3 code, whereby the mean of number of skilled occupations by NAICS code is computed using the Census data (following Minondo and Requena-Silvente, 2013). Our index of complexity is correlated with owner's education in the data.

[26]The definitions of these alternative proxies are as follows. For literacy, subjects were asked whether the sentences "The light balloon floated in the bright sky" and "A comfortable pillow is soft and rocky" make logical sense. For numeracy, subjects were presented with the following question: "Suppose you have K1,000 in a bank account with no bank fees. The bank pays interest of 10 percent each year. How much money will you have after 2 years?" Subjects were then presented with a choice of either "Less than K1,200", "K1,200 exactly", or "More than K1,200". For social skills, subjects stated how much they agreed, on a scale of 1-5, with the statement "I know what other people are feeling just by looking at them", as well as "Gaining happiness requires taking it away from others" and questions on how often the subject talks with others about social topics (e.g., sports) or personal topics (e.g., health).

different industries explains between one-half and three-fourths of the gender gap in sales and employment. The average level of sales and employees is the lowest in apparel manufacturing - where most women operate - and the highest in food manufacturing. The other sectors lie between these two. Controlling for sector, education, working time and marital status makes the coefficient on the female dummy insignificant for the sales gap in good weeks. This evidence is compatible with previous studies also finding that women enjoy less profitability and lower sales growth even controlling for extensive observable characteristics (Klapper and Parker, 2011; Campos et al., 2014; Hardy and Kagy, 2018). Recent evidence by the World Bank similarly finds that one-quarter of the gender gap in profits in the Democratic Republic of Congo can be explained by industry (Campos et al., 2019). As we have emphasized throughout this paper, one explanation for female industrial segregation is that when gender bias is large and institutions are weak, women cannot trust men.

2.4.3 Rule of law and legal institutions in Lusaka

Zambian rule of law is neither particularly good nor particularly bad for sub-Saharan Africa, but it is in the bottom half of countries worldwide. The country's score on the World Justice Project's Rule of Law Index is below Ghana and South Africa, but above Zimbabwe and Nigeria, and about the same as Russia and Mexico. Zambia's overall index of gender equality and gender equality in courts from the World Justice Project are among the lowest in Africa.

Few entrepreneurs in our sample made any use of Zambia's formal court system, which are notoriously slow and cumbersome. Instead, many of them rely on local adjudicators, known as "Market Chiefs," who exercise authority over transactions that occur within their own local market areas. There are approximately 80 formal markets in Lusaka. These fall under two broad categories: council and cooperative.

Cooperation appears to be easier within the market, and somewhat surprisingly, rents also appear to be lower, at least relative to space in well-travelled commercial thoroughfares. The offsetting downside of markets is that they are sometimes harder for customers to access and have shorter opening times.

The 30 council markets are regulated by Lusaka City Council and are led by a market officer (henceforth referred to as a chief) who is appointed by the Council. The appointed chief's functions are guided by legislation and market unit guidelines. Many markets also have a democratically elected chief, whose involvement in market affairs varies. The cooperative markets usually have a market committee of 6-10 members, including a democratically elected chair.

Figure 2.9 shows the spatial location of markets with squares. Markets generally have their own system of rules. They set fines for prohibited behaviour, such as drinking or insulting other marketeers on site, and rules of suspension for behaviour such as theft

or disobedience of the market committee. Most importantly, the market chiefs resolve disputes among marketeers and market customers.

One typical dispute was that a marketeer sold his plot within the market to two different people, taking money from both. The marketeer did not have assets to seize, beyond the plot, but the chief knew a lender had agreed to loan the marketeer some money. To solve the dispute, the chief ensured this loan money was given to one of the buyers.

In another dispute, a customer complained that a marketeer had failed to fulfill his promise to fix a television. The chief gave the marketeer a deadline of two days to meet the customer's request or pay a fine. As one chief said, "we want to handle disputes internally and peacefully first," and "only if this does not work do we go to the police."

Markets are an essential part of Lusaka's business environment: 30 percent of firms across all industries in our Census and 59 percent (1324) of manufacturers are located in formal markets.

To address the weaknesses of the formal court system, a Small Claims Court was founded in 2008 by an act of Parliament and opened in 2009. No legal representation is required, as the court is intended to hear minor cases, for example, relating to employment, borrowing and lending, insurance and fraud. The court only permits cases involving amounts up to 20,000KW (around 1,962 USD), which represents 150 percent of average sales in a good month.

To file a case with the Small Claims Court, a plaintiff must first produce a letter of demand which opens the case and serves notice to the defendant. This letter of demand brings with it a seal of the court and often by itself can be sufficient to recoup any contested amount (at a cost of 5Kw, around 50 cents in USD). Many cases get settled between the two parties at this initial stage. However, if the defendant refuses to pay the amount stated in the letter of demand, a court hearing is scheduled where the defendant has an opportunity to submit a written defense. Within a month of being served, the case is heard and, depending on the outcome, the losing party has to pay according to a certain schedule. If the plaintiff wins, the defendant has to pay back the money and also cover the legal costs.

If the Small Claims Court makes a decision on a case and the defendant does not settle the claim, a writ of execution is issued by the courts and bailiffs are engaged. The writ of execution is put on the file together with a report from the bailiffs about the property they seized, how much they realized at auction and whether the amount recovered covered the full claim. The bailiffs are not allowed to seize personal items such as clothing and the tools of trade of the individual. In practice, the plaintiff has to know the residence of the defendant in order to give detailed directions to the bailiffs (a sketch map is usually on file for these cases).[27]

[27] In cases where the amount realized from the seized property is not enough to cover the claim the court will tell the plaintiff to keep a look out if the defendant buys new property and contact the bailiffs so that they can seize the new property.

The time frame to settle the claim is usually outlined in the judgement by the commissioner. Defendants can be asked to pay in instalments over a period of time or to pay in 14 days; if they fail, bailiffs are engaged. In extreme cases where an individual is not able to pay, the person can be imprisoned. This outcome is rare because the amounts the court deals with are small and the defendant's relatives usually pitch in to help prevent the person being incarcerated.

The Small Claims Court currently has excess capacity, due to lack of awareness. In our data, only 21% of respondents had heard of it, and of those, 56% did not believe they had access. Market Chiefs are a far more common mechanism for resolving disputes. Anecdotally, Senior Clerks at the Small Claims Court note that women sue mostly men, especially their (former) employers. Figure 2.9 shows where the Small Claims Court is located in our zoom-in of the Census Map.

2.5 Trust, gender and institutions in Zambia: observational evidence

Partnerships in Lusaka resemble the world of incomplete contracts described by our theory. Written contracts are used only in one out of five partnerships. Collaborations are mostly triggered by time constraints (e.g., in fulfilling big orders), which limits entrepreneurs' ability to search for partners. In seventy-five percent of cases, partnerships form between entrepreneurs that are located close to each other. In this section, we first examine whether female entrepreneurs are less trusting and form fewer partnerships than their male counterparts. We then look at whether this gender gap is ameliorated by physical proximity to the two local institutions: Market Chiefs and the Small Claims Court.

2.5.1 Gender and trust

To examine whether women have a disadvantage in bargaining, we look at their trust levels and interactions with other businesses. Throughout the paper, we use the following survey measures of trust asked in the Manufacturers Survey and taken from the World Values Survey (WVS) and General Social Survey (GSS):[28]

- Trust GSS: Do you think that most people can be trusted or you cannot be too careful? (one-zero indicator variable)
- Trust Strangers: How much do you trust people you meet for the first time? (from 1 (not at all) to 4 (completely))

[28]We show validity checks of the survey measures of trust and trustworthiness using our experimental data. We find that our survey measures of trust are correlated with experimental trust-worthiness, as shown in previous research (Glaeser et al., 2000). Moreover, the survey cooperation measures are correlated with the number of tokens sent by the investors.

- Trust Neighbors: How much do you trust your neighbours? (from 1 (not at all) to 4 (completely))
- Trust in business: I am comfortable leaving my shop unattended during the day if I need to do something for 30 minutes (5-points Likert scale) [29]

These measures of trust can be interpreted in different, but related, ways. They can be seen as the internal psychic cost of betrayal, or as the individual belief in others' trustworthiness. That belief might reflect the reality that some people have a greater ability to enforce trustworthy behaviour relaying on either low-cost tools such as social sanctions, or high-cost tools, such as violence or courts. In the model, trustworthy behaviour reflects the existence of a high-cost enforcement tool (the court), but in reality, many forces may shape individual's answers to these questions. [30]

We couple these general trust questions with questions about business behaviour that involves trust, including the formation of business partnerships. We conducted extensive piloting to identify the most common cooperative activities that small-scale manufacturers engage in, across all industries. We then adapted the language used by our piloting participants to create the following four questions:

- Sometimes two or more businesses participate in a common order from a client, or one business subcontracts to other businesses part of an order. Have you ever done this with another business like yours? (share order)
- Sometimes businesses make joint orders of materials from suppliers. Have you ever done this with another business like yours? (joint buy)
- Sometimes businesses ask for advice (or give advice) to other firms doing their same activity, for instance on topics like: the production process, the market conditions, new technologies, business practices, suppliers. Have you ever done this with another business like yours? (advice)
- Sometimes businesses borrow (or lend) machines, materials or other assets from firms doing their same activity. Sometimes they hire (subcontract) employees who come from other firms doing their same activity for a short period of time. Have you ever done this with another business like yours? (lending)

In our analyses, we use both indicator variables that take on a value of one if an individual ever engaged in a particular activity and also an index of cooperative behaviour

[29] The Trust Neighbors and Trust Strangers variables were converted into dummy variables by combining low scores (1 and 2) and high scores (3 and 4). Low scores were given a value of 0, and high scores a value of 1.

[30] This flexibility allows us to interpret individual trust as a proxy of individual investment in social capital. As for any other form of capital, investing in social capital requires forming expectations on its returns, which can be affected by several determinants such as other's trustworthiness, risk aversion and the ability to punish cheating (Glaeser et al, 2000; Ashraf et al, 2009; Butler et al, 2010).

from their responses (averages of the four indicator variables).[31]We end this section by also discussing our limited information on the transfer of knowledge across Lusakan entrepreneurs.

Panel A of Table 2.6 shows the mean differences in trust between men and women. Women are less trusting across all three direct survey measures. Women are more likely to disagree with the statement that they would be comfortable leaving their shop unattended. Panel B of Table 2.6 shows that women are also less likely to engage in any of the four cooperative behaviours that we measure.

Table 2.C.2 shows that low levels of trust are associated with lower frequency of cooperation among entrepreneurs. These results support the view that the trust questions are capturing something real about the trustworthiness of the environment. Table 2.C.3 shows that cooperative behaviour is also correlated with our three measures of sales. While these correlations do not imply any causal relationship, they are compatible with the view that entrepreneurial activity benefits from the ability to form partnerships with others.

In the previous section, we documented that female entrepreneurs selected into less profitable industries. Our model suggests that this self-selection occurs because women choose industries with other women, because they are able to trust and partner with those women, either as customers or collaborators.[32] Figure 2.12 shows that women generally have fewer social interactions with other entrepreneurs in their industry, but that tendency disappears in industries that are not male dominated. Women who work in the disproportionately female industries have social interactions that are as frequent as men in those industries.

Perhaps the most important form of cooperation occurs when one urbanite shares knowledge with another. In these informal industries, many skills are often passed along from one worker to another. Yet chains of knowledge also can require trust. In many cases, a skilled worker who teaches a newcomer expects that student to serve as an apprentice. For centuries, the relationship between mentor and apprentice has been open to abuse. Apprentices, including Benjamin Franklin, run away to avoid promised service. Mentors typically gain the power to punish their students either physically or by harming their reputations, and that power also creates the potential for misuse, especially between a man and a woman.

Figure 2.11 shows that male entrepreneurs are far more likely to have been taught their trade by another entrepreneur or a family member. Female entrepreneurs are usually formally trained. Female entrepreneurs are also less likely to have taught others their trade, but this result is not robust to other controls.

Table 2.7 shows these results using a linear probability model with controls for business density and owner characteristics. Regression (1) shows that women are 19.1 percent

[31]Results are robust to alternative aggregations of these variables, such as a z-score.

[32]This latter explanation is in line with the results by Campos et al. (2014), who identify role model as one of the most important reasons for women to select into industries.

less likely to learn their trade from another owner. Regression (2) shows that they are 13.9 percent less likely to learn from another owner in the same industry. Regression (3) shows that they are 15 percent less likely to learn from family and friends, which seems more plausibly related to gender discrimination within the household than trust.

The last three regressions in the table show the relationship between the source of the owner's knowledge, sales and earnings. Regression (4) shows that in a good weak, women have sales that are .6 log points lower than men, but this difference is largely reduced if women have learned from another owner or family member. Regression (5) repeats this regression for sales in a bad week and finds quite similar results. Regression (6) repeats the regression using employment as the dependent variable. The pattern for employment is similar but the interaction between owner gender and source of skill is not significant. If human capital externalities rely on trust, then women may be unable to access those externalities and benefit fully from working in a dense urban environment.

2.5.2 Institutions and female trust

In this Section, we ask whether institutions can mitigate the gender gap in collaboration and earnings among Lusaka manufacturers. The model predicted that female entrepreneurship requires both female bargaining power and rule of law, and gender norms appear to be quite biased in Zambia. Yet conditional upon entry, the model predicts that women will be unequivocally more likely to partner with men when rule of law is higher, independent of gender discrimination.

The two conditions for partnership in Proposition 1 are $\frac{\pi-q-b}{\Delta+b} > \delta$ and $\frac{\pi-2q}{\Delta} > \delta$, depending on parameter values. Both conditions depend on the returns to partnership and the returns and costs of cheating, but as long as these parameters are held constant, the model strongly predicts that improvements in the quality of legal institutions will make partnership between men and women more likely.

We use two measures of institutional strength: physical proximity to the Small Claims Court and access to the justice offered by the chief of a formal market. We then compare the outcomes of female and male entrepreneurs who are located within a market, or close to the Small Claims Court, with their counterparts who are located either outside a market or further away from the Small Claims Court.

A primary question is whether these measures of access actually capture δ. If market chiefs or the Small Claims Court are thoroughly biased, then physical proximity to them will not engender cooperation by women. The closest mapping between these measures and our model occurs when women have no ability to enforce contracts against men without these institutions ($\delta = 1$), but that with these institutions δ rises to some higher number.

Table 2.8 and 2.9 show the following regression where cooperative behaviour is regressed on access to legal institutions and the interaction between these institutions and

gender. We run regressions of the form:

$$coop_{is} = \beta_1 Female_i + \beta_2 instProximity_i + \beta_3 Female_i * instProximity_i + X_i^{'}\delta_i + \epsilon_{is}$$

where $coop_{is}$ is the outcome variable for business i in sector s.

The dependent variables in columns (1)-(4) are the indicator variables that capture cooperative behaviour: whether the respondent said that they had lent/borrowed capital, given/received advice, participated in a common order from a client, or placed a joint order of materials with another business like their own. The dependent variable in column (5) is their mean. The set of controls X_i includes the following characteristics about the area around the business: the logarithm of the total number of businesses within 100 meters of business i, the logarithm of the total number of businesses in the same industry (NAICS3) within 100 meters around business i, and a dummy for whether the business is within 100 meters of a market ("business density controls"). Our empirical strategy compares the effect of institutional proximity for businesses of different genders that are exposed to similar demand conditions and agglomeration effects.

Table 2.8 shows that being in a market disproportionately increases the probability of sharing an order or giving advice for women. It also increases women's average cooperation.[33] This is consistent with the hypothesis that strong market leaders might provide support to women's interactions with other businesses, overcoming their disadvantage in bargaining power. Figure 2.10 shows that the increase in women's average cooperation in markets is not driven by markets where women's manufacturers are the majority, but also happens in markets where men's manufacturers are the majoritarian group.

Table 2.9 shows the relationship between cooperation and the distance to the Small Claims Court. Distance from the Small Claims Court disproportionately reduces cooperation for women.[34] Figure 2.13 shows the proportion of people of each gender cooperating with other businesses based on distance from the Small Claims Court. Cooperation is stable for men, with the exception of the more distant areas. If women's business location is farther from the Small Claims Court, then cooperation diminishes. Both tables tell a consistent story in which access to legal institutions disproportionately enables cooperation by female entrepreneurs.

Tables 2.10 and 2.11 turn to sales, which is our primary measure of economic success. Table 2.10 regresses three measures of sales on a dummy indicating the business-owner's

[33]This relationship remains the same even when including business owner controls (such as whether the owner keeps written records of purchases made, age of the business owner, owner's marital status, among others). It is also qualitatively the same when controlling for industry fixed effects, but the direct effect of gender is attenuated. See table 2.C.4 of the appendix.

[34]This relationship remains the same even when including business owner controls (such as whether the owner keeps written records of purchases made, age of the business owner, owner's marital status, among others). It is also qualitatively the same when controlling for industry fixed effects, but the direct effect of gender is attenuated. See table 2.C.5 of the appendix.

gender, a dummy on whether the business is located within a market, and their interaction. All regressions include industry fixed-effects, business density controls and business owner controls.[35] Women sell less than men in all three specifications and in two of the specifications the difference is significant.

The interaction term is significant, suggesting that female businesses particularly benefit from locating within a market and having access to a market chief. Given the central role of the Chief in dispute resolution, this result may reflect women's increased ability to trust when there is a strong enforcing institution such as the markets chief.

In Table 2.11, we perform exactly the same analysis as in Table 2.10, but now focusing on the Small Claims Court.[36] There seems to be no strong evidence on the Small Claims Court significantly affecting businesses' sales, neither for the female nor for the male-owned businesses. Indicator variables for the distance to the Small Claims Court, and their interaction with the gender indicator, are largely insignificant. One natural interpretation of these results is that the Market Chief is a far better known and more effective remediation mechanism than the Small Claims Court, which is largely unknown in our sample.

Despite the robustness of these trends to different controls, unobservable characteristics of entrepreneurs located into markets might be driving our results. A particular worry is that the higher density of businesses in markets might mechanically increase the opportunities - and thus the incentives - for cooperation. We address this issue in the next section, where we present results from a lab-in-the-field experiment that shows the causal effect of institutions on trust and business cooperation.

2.6 Trust, gender and institutions in Zambia: experimental evidence

To provide causal evidence of the impact of institutions on trust and business cooperation, we perform an embedded experiment with a sample of Lusaka entrepreneurs. These experiments, an adapted version of the trust/investment game pioneered by Berg et al. (1995), examine whether institutional support particularly impacts female players. We run the games with actual entrepreneurs located inside or around formal markets and involve actual local institutions: Market Chiefs and the Small Claims Court judges.

[35]Owner controls include whether the business owner keeps written business records of every purchase and sale made, whether her/his business is registered individually, whether the business owners trust their neighbours, how old the business is, how many days the business owner spends working in the business, age of business owner and whether business owner is married or not.

[36]This relationship remains nearly the same when including industry fixed-effects and business owner controls. See table 2.11 of the appendix.

2.6.1 Experimental design

Using data from the Census, we randomly selected participants for the experiment, stratifying by gender and whether the business is located within a market. In order to avoid imposing on marketeers' time and to increase control, we conducted the game within people's shops. Surveyors communicated with each other electronically to convey to a player the choice of their partner in real time.

Using the Lusaka Census of Urban Entrepreneurs, we constructed lists of all entrepreneurs within each market as well as all entrepreneurs within 500 meters of the market border. The sample selection procedure for the experiment was designed to oversample manufacturers and women in a selection of Lusaka's 80 dense urban marketplaces. We chose sixteen markets for study that contained the largest numbers of manufacturers. For each market, we used the 2016 Census data to assemble a list of entrepreneurs located inside the market and a list of entrepreneurs outside the market, but within 500 meters of its borders.

We oversampled the manufacturing sector by placing all of the manufacturers at the top of the list in random order and all other businesses at the bottom of the list in random order. If a marketeer could not be located or refused to participate, we replaced the marketeer with the next female on the list. As soon as all women had been exhausted, the recruiters simply moved to the next marketeer on the list. The survey team visited businesses on the lists in this way until they had valid responses for 24 businesses inside each market and 6 businesses outside each market. We recruited a total of 480 participants across the 16 marketplaces.[37]

The experiment consists of a modified version of the trust game (Berg et al, 1995), framed as an opportunity to invest in another person's business opportunity. Player A (the Investor) was given 10 tokens that could either be kept or invested in the business of Player B (the Trustee). The Trustee received three times the number of invested tokens and must decide how many to return to the Investor. The Trustee used the strategy method: before receiving the Investor's tokens, he or she completed a matrix indicating how many tokens they would like to return for a given amount sent. The decision in their matrix was followed even if they wanted to change it after the Investor's offer.

Participants played the game for two rounds in total, but they switched roles between the first and the second round. Players who started playing as an Investor (Trustee) would play as a Trustee (Investor) in the second round. After the game was complete, players could exchange tokens for real money. The sessions were on average 90 minutes long. Investors earned on average 11.4KW (1 dollar at the time) and Trustees earned 6.6KW (0.60 dollars) per round. These amount to 3.6 percent and 2.1 percent of average daily earnings.[38] In addition, all players received a participation fee of 35KW (3.5 dollars)

[37] 475 of the games responses were usable in the data analysis.

[38] In 7 percent of chief-treated games and 9 percent of court-treated games, the Investor asked for arbitration. Average earnings for all games before arbitration were 11.3 tokens for the Investor and 6.7 for the Trustee.

as a token of appreciation for their time.

We cross-randomized two experimental conditions (table 2.12). The first condition tests the effect of institutions on trust. We randomly assigned pairs of players to three groups: a control group that receives no access to institutions, a first treatment group that has access to the Small Claims Court and a second treatment group that has access to the market chief. If the Investor is dissatisfied with the number of tokens received, then the Investor may ask the experimenters to go on their behalf to the court or chief to adjudicate. The invitation follows:

> If you think that the number of tokens sent back by Player B is not fair, you can ask us to call the chief (senior clerk at the Small Claims Court) on your behalf, to decide how many tokens each of you should get. The chief (senior clerk) will then decide how many tokens each of you should get. The chief's (senior clerk's) ruling is based only on your choices and the choices of player B, and we will tell you their decision accordingly. The chief (senior clerk) does not know anything about you and the other player, only your choices. Player B will know that you can complain to the chief (senior clerk). In the case that you complain, the final division of tokens will be determined by the ruling of the chief (Small Claims Court).

> Player A can ask us to consult the chief (senior legal clerk at the Small Claims Court) on his/her behalf, to decide how many tokens each of you should get. The chief's (senior clerk's) ruling is based only on your choices and the choices of player A, and we will tell you their decision accordingly. The chief (senior clerk) does not know anything about you and the other player, only your choices. In the case that player A complains, the final division of tokens will be determined by the ruling of the chief (Small Claims Court).

The implementation of the complaint required us to ensure that participants trusted that the experimenters would call the institution to complain, and that the Chief and the Small Claims Court Senior Clerk would understand the game and act as if it were a real-life dispute. We provide details on our solutions to these challenges in Appendix B.

In previous sections, we showed that cooperation levels are higher for businesses located inside formal markets than outside formal markets. The second experimental condition allows us to explore whether this result is driven by in-group vs out-group dynamics which could increase cooperation within markets independently of contract enforcement (Kranton et al., 2018). Pairs of players were randomly assigned to two groups. For the first treatment group, both players were drawn from within the same market. For the second treatment group, one participant was drawn from the market, while the other from outside the market. For all groups, the players are told whether they are playing

with a person from the same market or an outsider. The two experimental conditions were cross-randomized, yielding five distinct experimental groups.

In order to assure understanding of the game, everyone had to correctly answer understanding checks about the rules of the games with the surveyors before proceeding.

We also implemented a Pre-Games Survey with questions about access to institutions and exposure to crime. We asked entrepreneurs whether they had had disputes with other business owners, how disputes had been resolved, and whether respondents had heard of the Small Claims Court. We also presented business owners with a hypothetical scenario in which one marketeer failed to pay back a loan to another. Business owners were asked whether they thought that the market chief, Small Claims Court, and police would be fair and/or slow in arbitrating the dispute.[39]

Table 2.C.6 shows mean differences by gender of these variables. Even though men and women are equally likely to have had work-related disputes and to know about institutions such as the Small Claims Court, men are significantly more likely to seek out another person, group or institution for help in resolving the dispute. This suggests women believe that institutions will not be effective for them. The fact that both women and men are similarly prone to be victims of theft, but the thief is rarely caught when the case involves female entrepreneurs, suggests why women may hold such a belief. Men and women are equally pessimistic about the fairness of police or the chief on average, but women are more likely than men to think that either the market chief or the police are slow in dispute resolution. Women reported being in general less trusting, feeling less comfortable when leaving their shops unattended. Overall, Table 2.C.6 indicates that institutions may be more effective for men, which leads women to rely less on these institutions and to have lower trust levels. Yet even biased institutions may be better for women than a complete absence of rule of law.

The pre-games survey showed a significant difference between market chiefs and the Small Claims Court. Market chiefs are a well-known institution to both men and women. For instance, one participant said that, in his market, the marketeers "have resolved to take all disputes to the market chairman". Half of the games participants think that the chief is "usually or always" fair in solving disputes.

The participants also note the flaws of Market Chiefs. One-fifth of men and one-quarter of women think that the chief is "usually or always" slow in solving disputes. Moreover, across markets, the chief's perceived fairness decreases as the proportion of female manufacturers increases. In the games, we improve upon this well-known institution by ensuring anonymity, which should eliminate any possibility of gender bias, and by bringing claims to the Chief on the participants behalf.

The subjects' awareness of the Small Claims Court's existence was far more limited. Eighty percent of our sample had not heard of the Small Claims Court before. Among the

[39]The pre-game survey might have primed participants to think about their past experiences with institutions before the play. This should not be an issue for the interpretation of the results as long as this effect is the same across experimental conditions.

20 percent who had heard of the Small Claims Court, half gave a negative answer to the question "whether they thought that business owners like them had access to the SCC". Five women out of 25 (and 7 men out of 76) mention that the SCC has high costs and "terribly" long procedures.[40] This limited and faulty knowledge, which we discovered only after setting the experiment in motion, led us to conclude that the Small Claims Court was not understood as an institution and we therefore present our results on the Small Claims Court only in Appendix tables.

We will interpret having access to the market chief - as compared to the control - as an improvement to the local institutional quality.

2.6.2 Results on trust and trustworthiness

Table 15 shows our primary results. Regression (1) uses the amount sent in the trust game as our experimental measure of trust. In this regression, we find that women without access to the Market Chief trust much less than men. This trust gap is compatible with all of the previous results in this paper showing the female entrepreneurs collaborate less then man.

This trust difference is almost completely eliminated when we introduce the market. As Figure 2.14 shows, men and women have a sizable gap in trust without the Market Chief. With the market chief, this gap disappears. This result suggests that experimentally generated differences in access to rule of law appears to encourage cooperation disproportionately for women in a developing world setting.

Regression (2) examines trustworthiness using the average return ratio, as in Glaeser et al., (2000). We divide the number of tokens the Trustee would return by the number of tokens available (return ratio) for each possible amount of tokens received, and then average over all return ratios. Somewhat surprisingly, women actually return less than men. The mean level of trustworthiness does not increase when we introduce the Market Chief, although Figure 2.14 shows that there is a mild reduction in the share of very low return ratios with the Market Chief.

Why did the Market Chief increase trust, especially for women, but not trustworthiness? This apparent puzzle reflects one major difference between our experiment and reality. The Market Chief in the experiment has no ability to punish, just to change the return ratio. Consequently, there was no real reason for Trustees to alter their behaviour, since it was unclear what the Market Chief would do. However, the Investors could still feel more protected, because even if the Trustee behaved badly, the Market Chief could still protect their investment. Regressions (3) and (4) shows the final earnings. Regression (3) shows that female and male Investors in the control group tend to receive the same number of tokens at the end of the game. The chief raises the Investors' total earnings,

[40] For instance, one female participant felt like the SCC "is a waste of time due to complicated procedures and costs. Also most people are illiterate and don't understand the services provided as well as what it's for".

with the effect being statistically the same for entrepreneurs of both genders. Regression (4) shows that the chief has little impact on the earnings of the Trustees.[41]

2.7 Concluding remarks

The developing world's rapid urbanization generates economic opportunities because urban density allows for collaboration and trade that is just not possible on far-flung farms. Yet the ability to interact positively is limited if one partner consistently fears expropriation by the other. Consequently, rule of law complements urbanization, not just because effective public order reduces activities that generate negative externalities, like water pollution, but also because rule of law enables a wider range of positive interactions.

The advantages of rule of law are particularly high when one set of actors has more power in a state of anarchy than the other. If men are more prone to use violence than women, then this will reduce women's bargaining power without legal support, and may cause inter-gender partnerships to break down or fail to form. In Lusaka, we believe that that industrial segregation of female entrepreneurs partially reflects the advantages of being able to trade within genders.

A central theme of this paper, and an implication of our cross-sectional and experimental work, is that even gender-neutral applications of rule of law can have benefits that accrue disproportionately to women. When rule of law is absent, the threat of violence looms over even seemingly innocuous market transactions. Men have displayed a disproportionate propensity towards violence across almost every known human society (Wilson and Herrnstein, 1985).Consequently, enforcing basic rule of law can make it easier for women to transact with men.

A second theme is that female entrepreneurship can be blocked both by weak legal institutions and by social norms that favour men and male bargaining power. Women can either lose in ex ante bargaining or ex post expropriation. Women will only enter entrepreneurial fields where much of the benefits require dealing with men, when they are protected from both types of loss.

[41] See the Appendix for results of the game splitting the sample by all the different institutional treatments.

2.8 References

Acevedo, P., Cruces, G., Gertler, P. & Martínez, S. (2018). *Living Up to Expectations: How Vocational Education Made Females Better Off but Left Males Behind.* Working paper.

Algan, Y., & Cahuc, P. (2010). Inherited Trust and Growth. *American Economic Review*, 100(5), 2060-92.

Alibhai, S., Buehren, N., Papineni, S., & Pierotti, R. (2017). *Crossovers: female entrepreneurs who enter male sectors - evidence from Ethiopia*. Policy Research working paper WPS 8065. Washington, D.C.: World Bank Group.

Attanasio, O., Kugler, A. & Meghir, C. (2011). Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 3(3), 188-220.

Bachas, P., Gertler, P., Higgins, S., & Seira, E. (2017). *How debit cards enable the poor to save more*. NBER working paper w23252. National Bureau of Economic Research.

Behrer, A.P., Glaeser, E.L., Ponzetto, G.A.M., & Shleifer, A. (2019). *Securing Property Rights*, CREI Working Papers.

Berg, J., Dickhaut, F., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10(1), 122-142.

Berge, L., Bjorvatn, K., & Tungodden, B. (2015).Human and Financial Capital for Microenterprise Development: Evidence from a Field and Lab Experiment. *Management Science*, 61(4), 707-722.

Calderon, G., Iacovone, L., & Juarez, L. (2017). Opportunity versus Necessity: Understanding the Heterogeneity of Female Micro-Entrepreneurs, *World Bank Economic Review*, 30, 86-96.

Campos, F., Coleman, R., Conconi, A., Donald, A., Gassier, M., Goldstein, M., Chavez, Z., Mikulski, J., Milazzo, A., Paryavi, M., Pierotti, R., O'Sullivan, M., & Vaillant, J. (2019). *Profiting from Parity: Unlocking the Potential of Women's Businesses in Africa. Main Report (English)*. Washington, D.C.: World Bank Group.

Campos, F., Goldstein, M., McGorman, L., Munoz Boudet, A.M., & Pimhidzai, O. (2014). Breaking the metal ceiling: female entrepreneurs who succeed in male-dominated sectors in Uganda. *Africa region gender practice policy brief 9*. Washington, DC: World Bank Group.

De Mel, S., McKenzie, D. & Woodruff, C. (2014). Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. *Journal of Development Economics*, 106, 199-210.

De Mel, S., McKenzie, D., & Woodruff, C. (2008). Returns to Capital in Microenterprises: Evidence from a Field Experiment. *The Quarterly Journal of Economics*, 123(4), 1329-1372.

De Mel, S., McKenzie, D., & Woodruff, C. (2009). Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns. *American Economic Journal: Applied Economics*, 1(3), 1-32.

Fairlie, R. W., Morelix, A., & Tareque, I. (2017). *2017 Kauffman Index of Startup Activity: National Trends.* Ewing Marion Kauffman Foundation.

Glaeser, E., Laibson, D., Scheinkman, J., & Soutter, C. (2000). Measuring trust. *Quarterly Journal of Economics*, 115(3), 811-846.

Global Partnership for Financial Inclusion & International Finance Corporation. (2011). *Strengthening Access to Finance for Women-owned SMEs in Developing Countries*. Washington, DC: Global Partnership for Financial Inclusion and International Finance Corporation.

Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic perspectives*, 20(2), 23-48.

Hardy, M., & Kagy, G. (2018). Mind the (Profit) Gap: Why Are Female Enterprise Owners Earning Less than Men? *AEA Papers and Proceedings*, 108, 252–255.

Hsieh, C., & Klenow, P. (2010). Development Accounting. *American Economic Journal: Macroeconomics*, 2(1), 207-223.

Kelley, D.J., Baumer, B.S., Brush, C., Greene, P.G., Mahdavi, M., Majbouri, M., Cole, M., Dean, M. & Heavlow, R. (2017). *Women's Entrepreneurship 2016/2017 Report.* Global Entrepreneurship Research Association, London Business School.

Klapper, L. F., & Parker, S. C. (2011). Gender and the Business Environment for New Firm Creation. *World Bank Research Observer*, 26(2), 237-257.

Knack, S., & Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics*, 112(4), 1251-1288.

Kranton, R., Pease, M., Sanders, S., & Huettel, S. (2018). *Groupy and non-groupy behavior: Deconstructing bias in social preferences*. Working Paper, Duke University, Durham NC.

Marshall, A. (1890). *Principles of Economics*. London: Macmillan.

Minondo, A., & Requena-Silvente, F. (2013). Does complexity explain the structure of trade?. *Canadian Journal of Economics*, 46(3), 928-955.

Morelix, A., Fairlie, R. W., Reedy, E. J., & Russell-Fritch, J. (2016). *Kauffman Index of Startup Activity: State Trends 2016.* Ewing Marion Kauffman Foundation.

OECD (2012). Gender Equality in Entrepreneurship. In *Closing the gender gap: Act now*, (part IV, 273-343). OECD Publishing.

Shah, M. (2012). *The Informal Sector in Zambia: Can it Disappear? Should it Disappear?* International Growth Centre Working Paper.

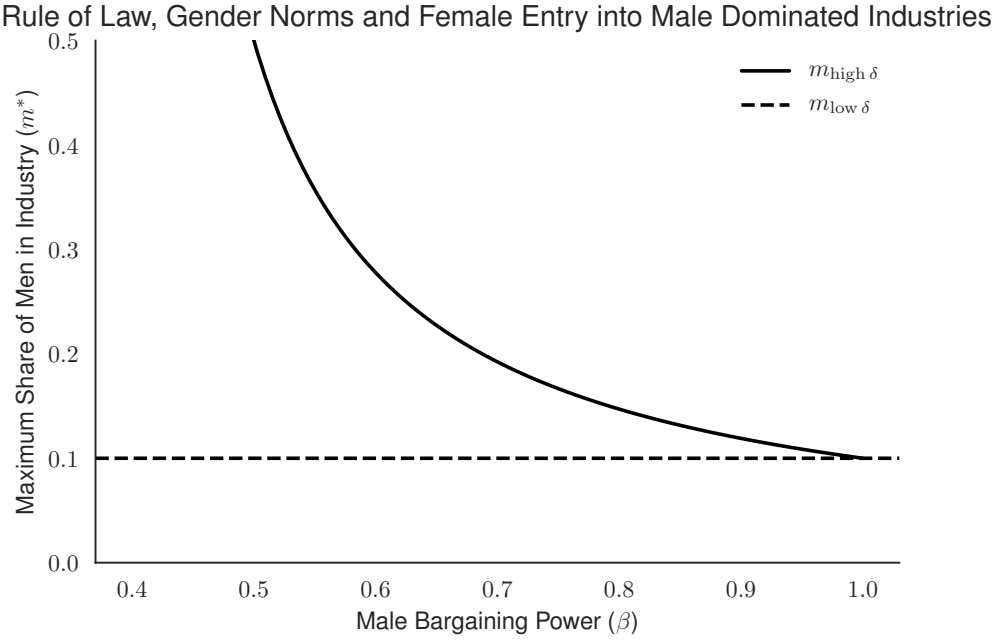
Singer S., Herrington, M., & Menipaz E. (2018). *Global Entrepreneurship Monitor 2017/18*. Global Entrepreneurship Research Association, London Business School.

Wilson, R.J.Q., & Herrnstein, R. (1985). *Crime and Human Nature: The Definitive Study of the Causes of Crime*. New York: Simon & Schuster.

Xavier, S. R., Kelley, J. K., Herrington, M. & and Vorderwülbecke A. (2012). *Global Entrepreneurship Monitor 2012*. Global Entrepreneurship Research Association, London Business School.

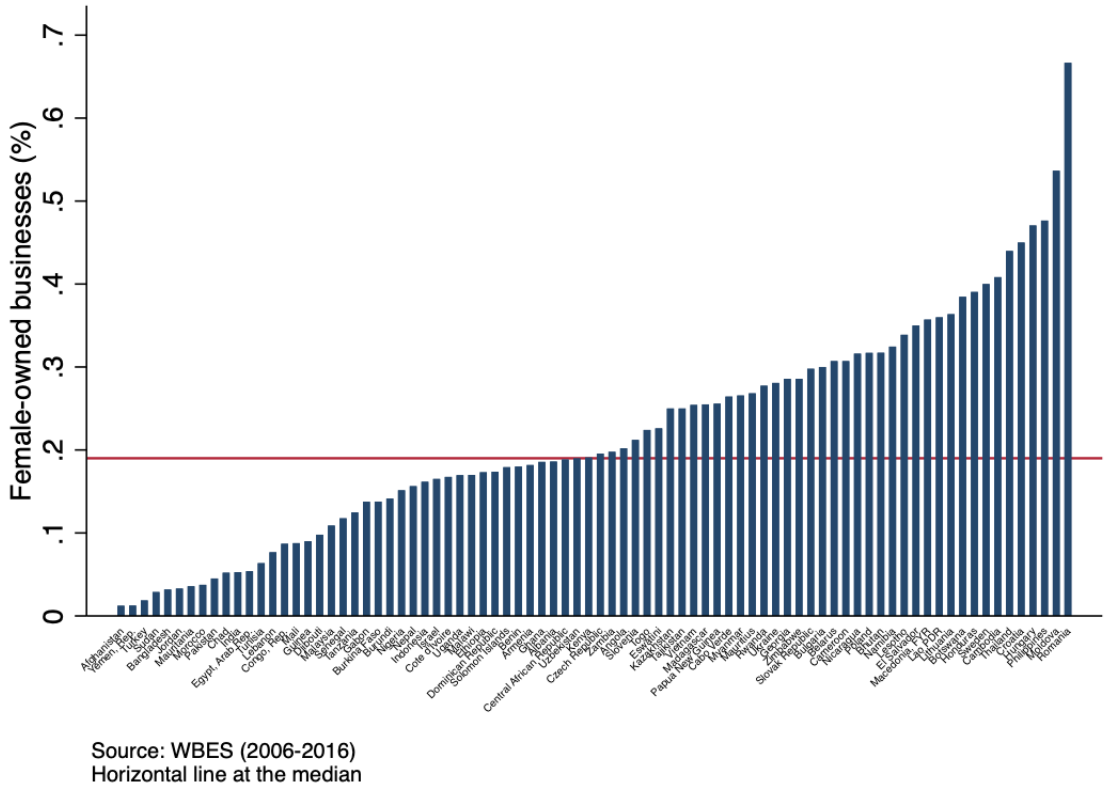
2.9 Figures

FIGURE 2.1: Theoretical predictions



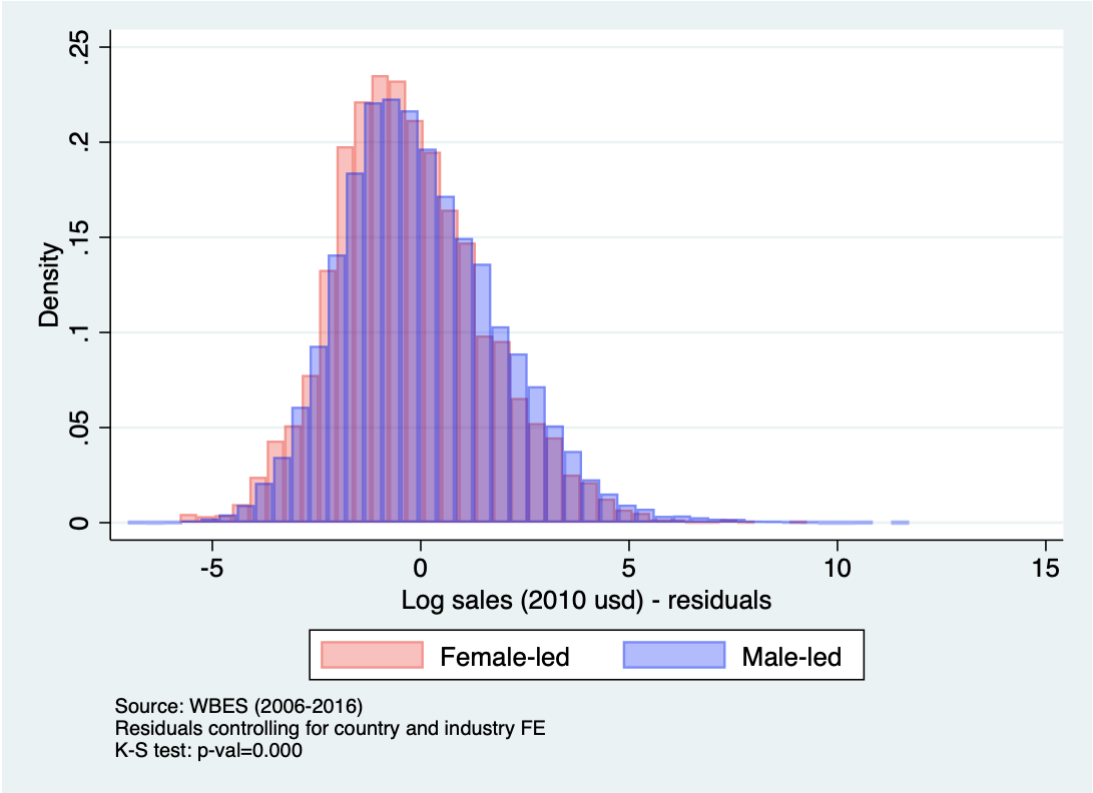
Note. The two lines both capture the maximum share of men in an industry that women will enter. The bottom line shows the case where rule of law is weak and δ is high enough so that women will not partner with men. In that case, the maximum share is $1 - \theta$ which we assume to be .1. The top line is decreasing with β and it shows the case where δ is low enough so that women will partner with men, and $\frac{\delta\Delta}{\pi - 2q}$ is assumed to be .2. The threshold for entry is reasonably high when δ and β is low, but if either δ or β are high, then women will not enter into male-dominated fields.

FIGURE 2.2: Female-owned firms across countries



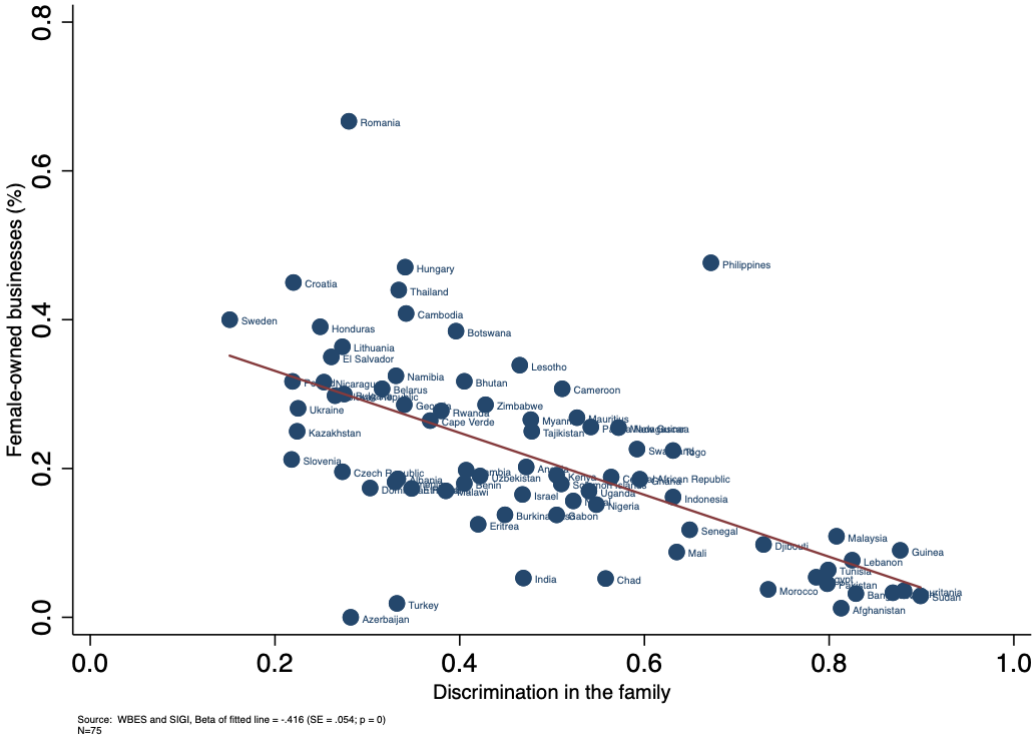
Note. This Figure shows the proportion of female-led businesses across countries. Female-led firms are defined as firms with a weak majority of female owners. We consider the most recent year of the World Bank Enterprise Survey for each country (years from 2009 to 2016). We limit the sample to businesses which are sole proprietorship or partnerships. The final sample excludes countries with less than 10 observations in those categories (5%). The number of countries in this figure is 84.

FIGURE 2.3: Business earnings by gender of the owner



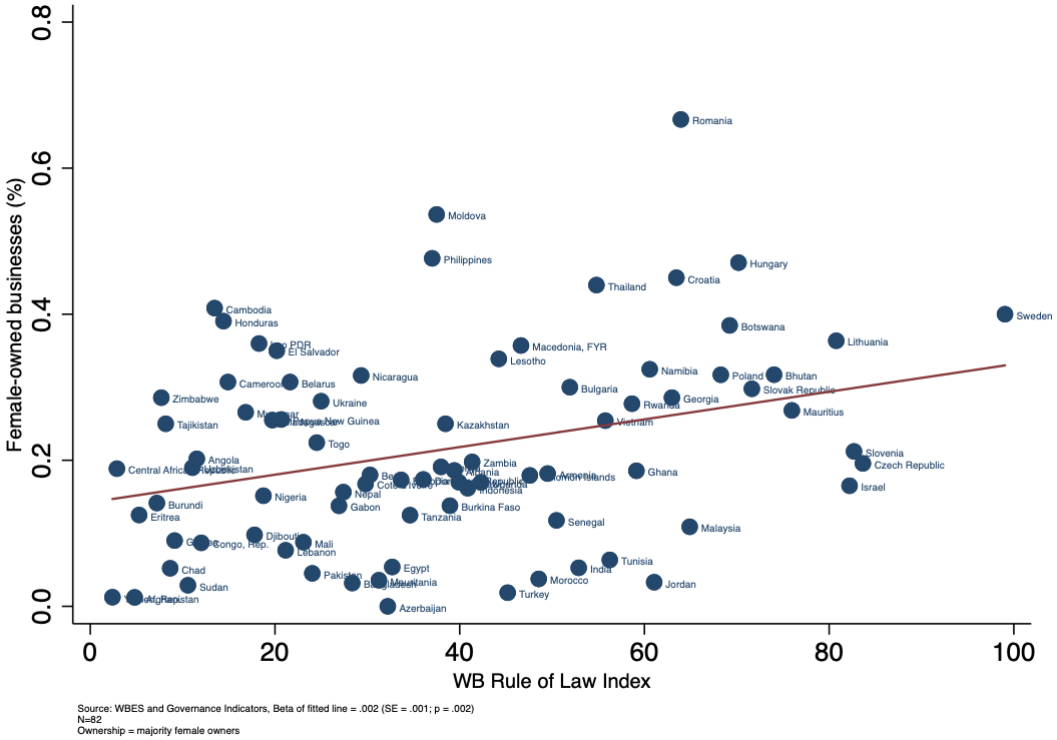
Note. This Figure shows the gender-specific distributions of logsales in 2010 USD, after controlling for country and industry fixed effect.We consider the most recent year of the World Bank Enterprise Survey for each country (years from 2009 to 2016). We limit the sample to businesses which are sole proprietorship or partnerships. The final sample excludes countries with less than 10 observations in those categories (5%). Female-led businesses are defined as firms with a weak majority of female owners. The number of countries in this figure is 82.

FIGURE 2.4: Female business ownership and family discrimination



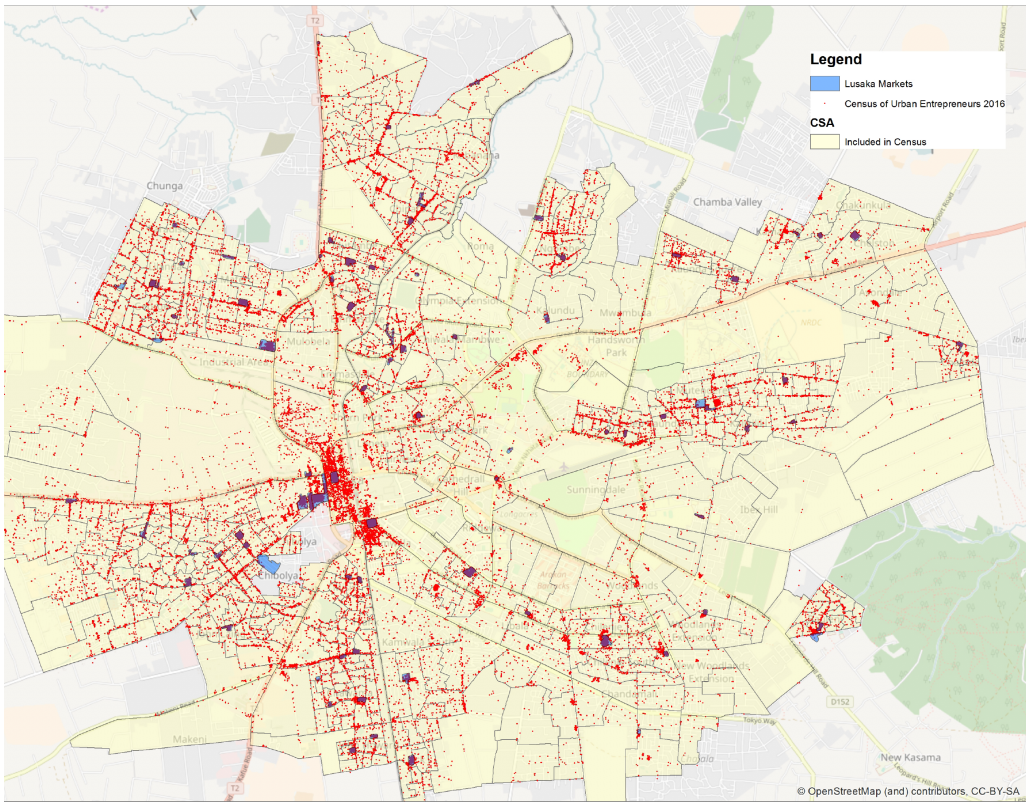
Note. This figure shows the correlation between a country's percentage of female-owned businesses and the SIGI measure of family discrimination. Female-led firms are defined as firms with a weak majority of female owners. We consider the most recent year of the World Bank Enterprise Survey for each country (years from 2009 to 2016). We limit the sample to businesses which are sole proprietorship or partnerships. The final sample excludes countries with less than 10 observations in those categories (5%). The variable for family discrimination is constructed as the weighted average of the following subcomponents: laws on child marriage, household responsibilities, inheritance laws and divorce laws. We use the SIGI 2019. The number of countries with both variables available is 75.

FIGURE 2.5: Female business ownership and rule of law



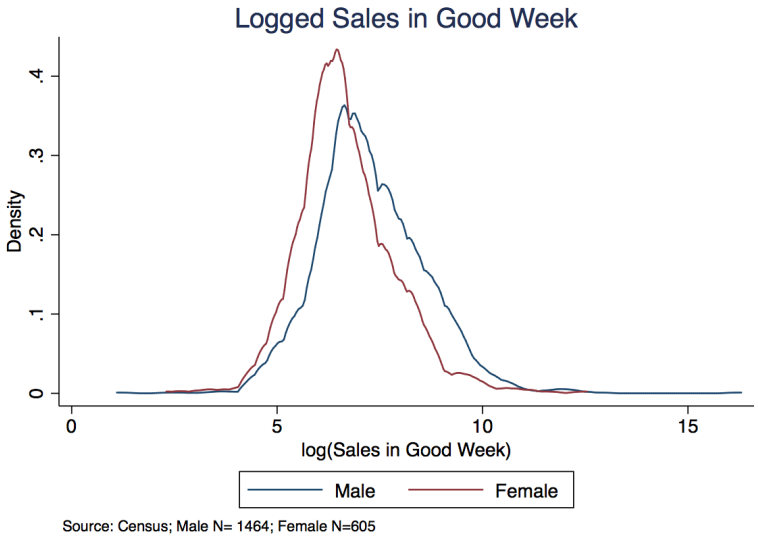
Note. This Figure shows the correlation between a country's quality of contract enforcement as measured by the World Bank Worldwide Governance Indicators (Rule of Law Rank Percentile) and its percentage of female-owned business from the WBES. Female-owned firms are defined as firms with a weak majority of female owners. We consider the most recent year of the World Bank Enterprise Survey for each country (years from 2009 to 2016). We limit the sample to businesses which are sole proprietorship or partnerships. The final sample excludes countries with less than 10 observations in those categories (5%). We use the 2017 Rule of Law Rank Percentile from the World Bank Worldwide Governance Indicators. The number of countries with both variables available is 82.

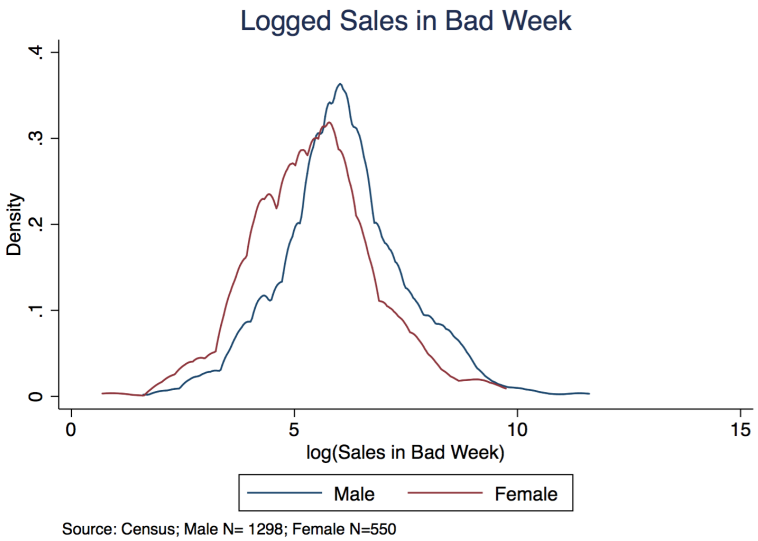
FIGURE 2.6: Census coverage



Note. This figure shows the spatial distribution of businesses in the Lusaka Census of Urban Entrepreneurs. It covers 48,163 establishments in Lusaka. The blue squares show where markets exist.

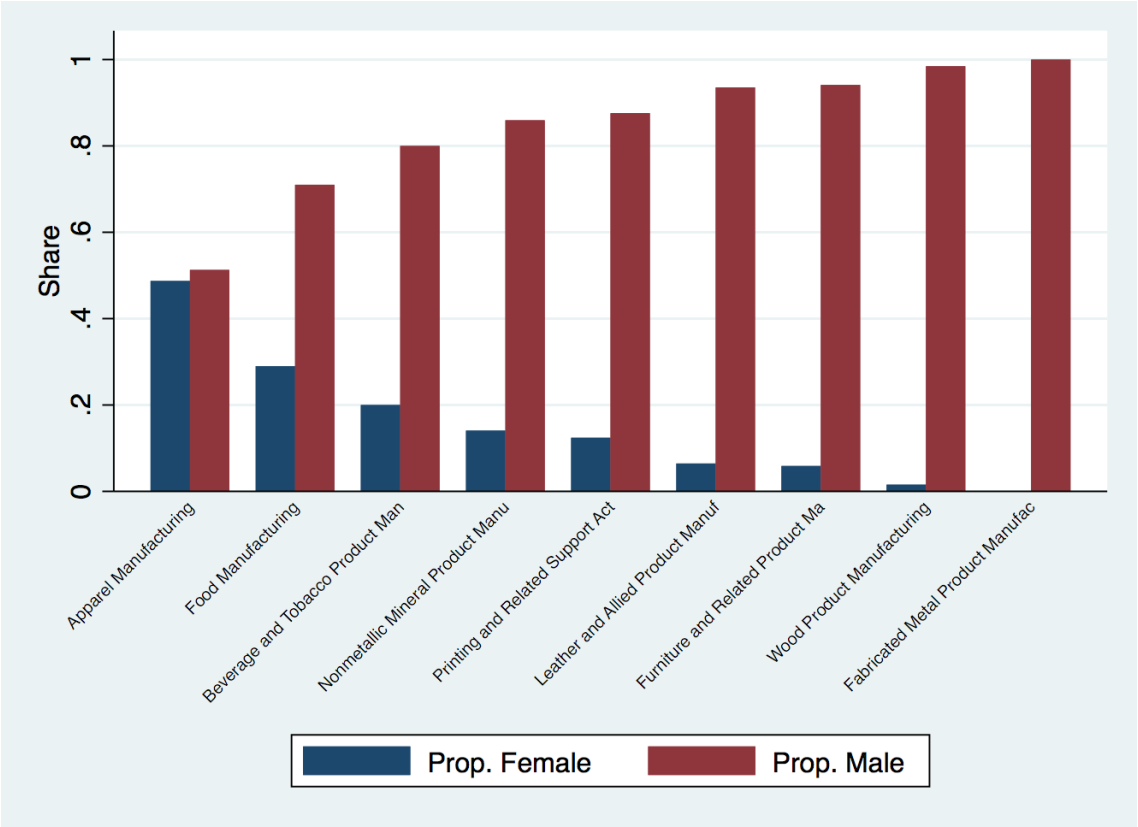
FIGURE 2.7: Sales in good and bad weeks by gender





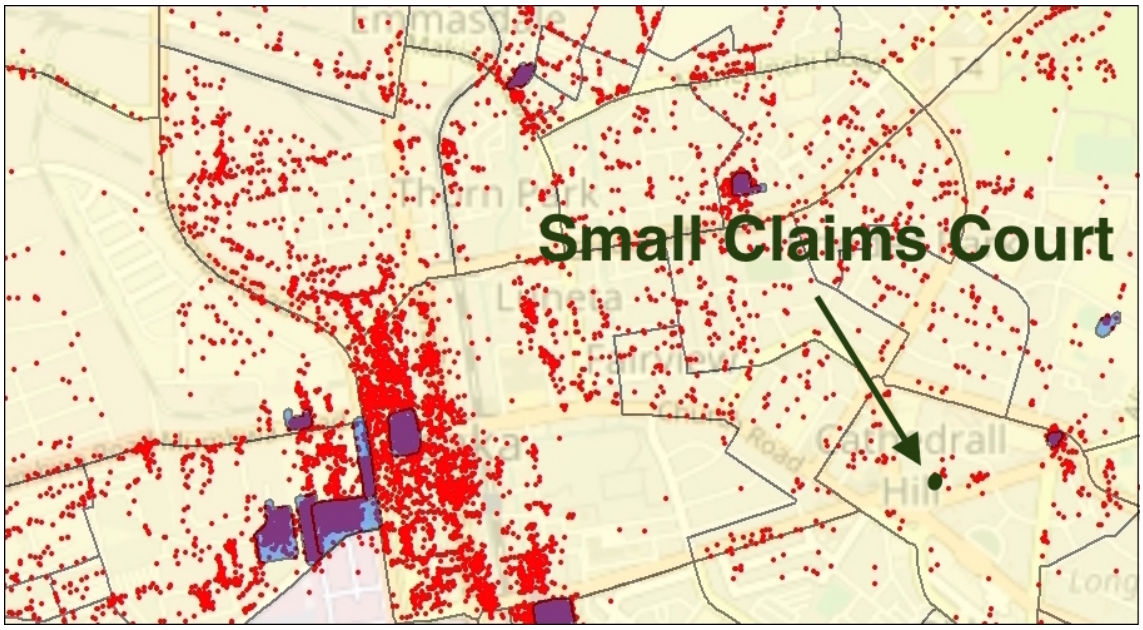
Note. This figure shows the kernel density of logged sales in good and bad weeks by gender. The p-value of the Kolmogorov-Smirnov equality of distributions test is 0.00. The p-value of the variance test ratio is 0.001 for sales in good weeks and 0.40 for sales in bad weeks. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

FIGURE 2.8: Distribution across industries by gender



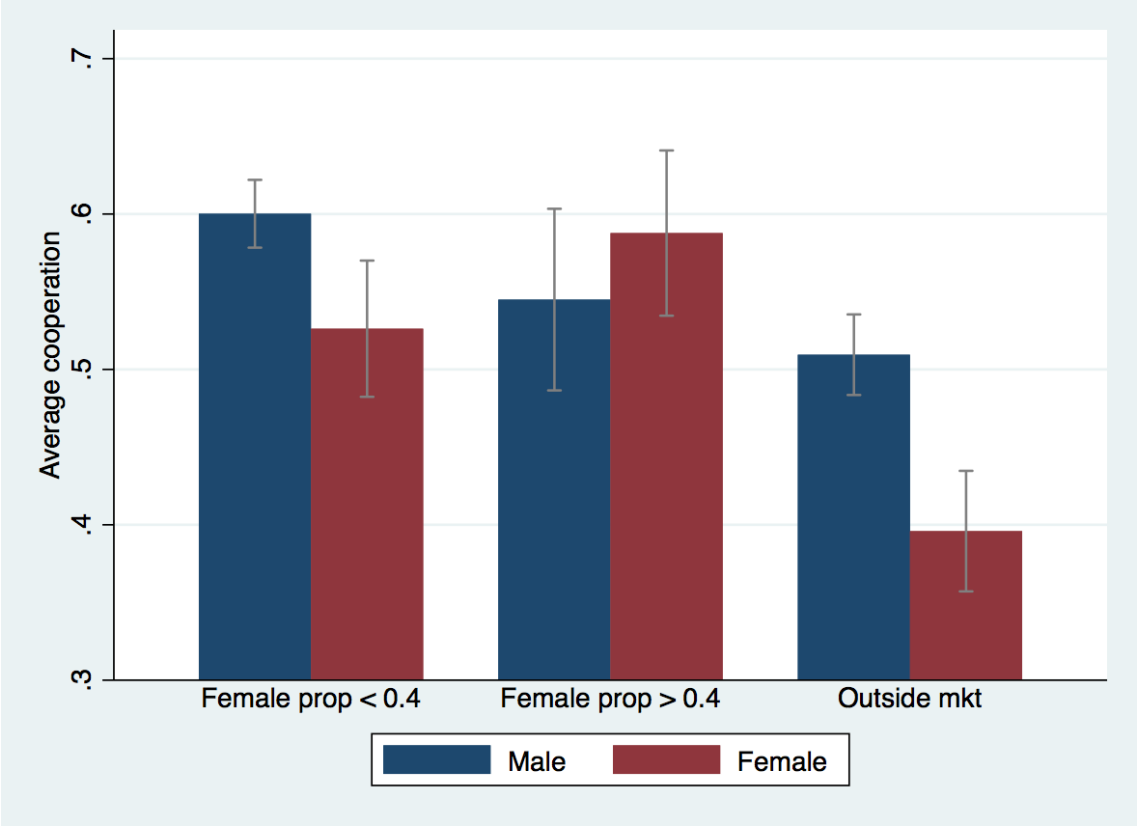
Note. This figure shows the industries in which manufacturers from the Lusaka Census of Urban Entrepreneurs operate, by gender. Industries are ordered by female shares.

FIGURE 2.9: Small Claims Court location



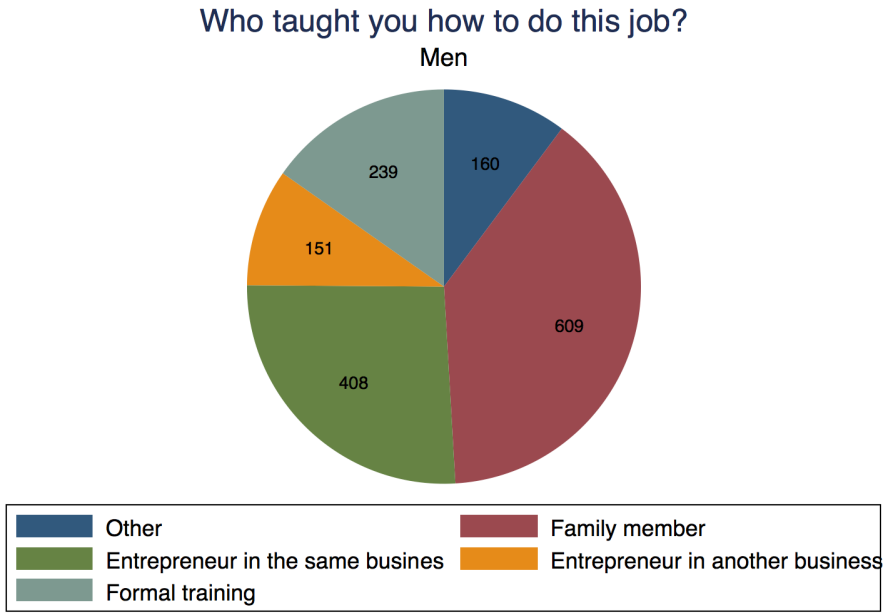
Note. This figure shows a zoom-in of the spatial distribution of businesses in the Lusaka Census of Urban Entrepreneurs. The blue squares show where markets exist and the arrow the Small Claims Court.

FIGURE 2.10: Cooperation by gender and market location

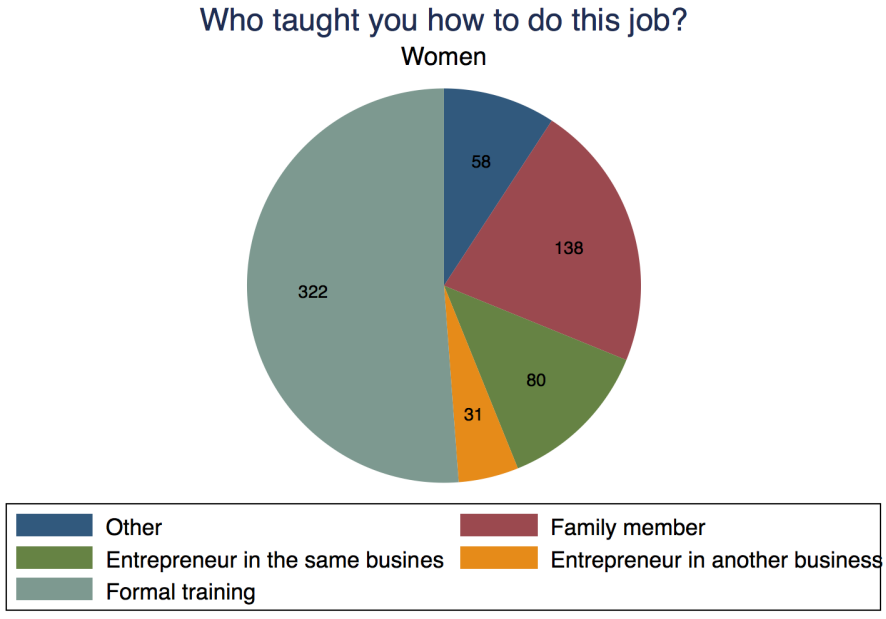


Note. This figure shows the average cooperation of female and male-led businesses in markets with a large or a small share of female businesses, or outside of markets. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

FIGURE 2.11: Who did entrepreneurs learn their trade from?



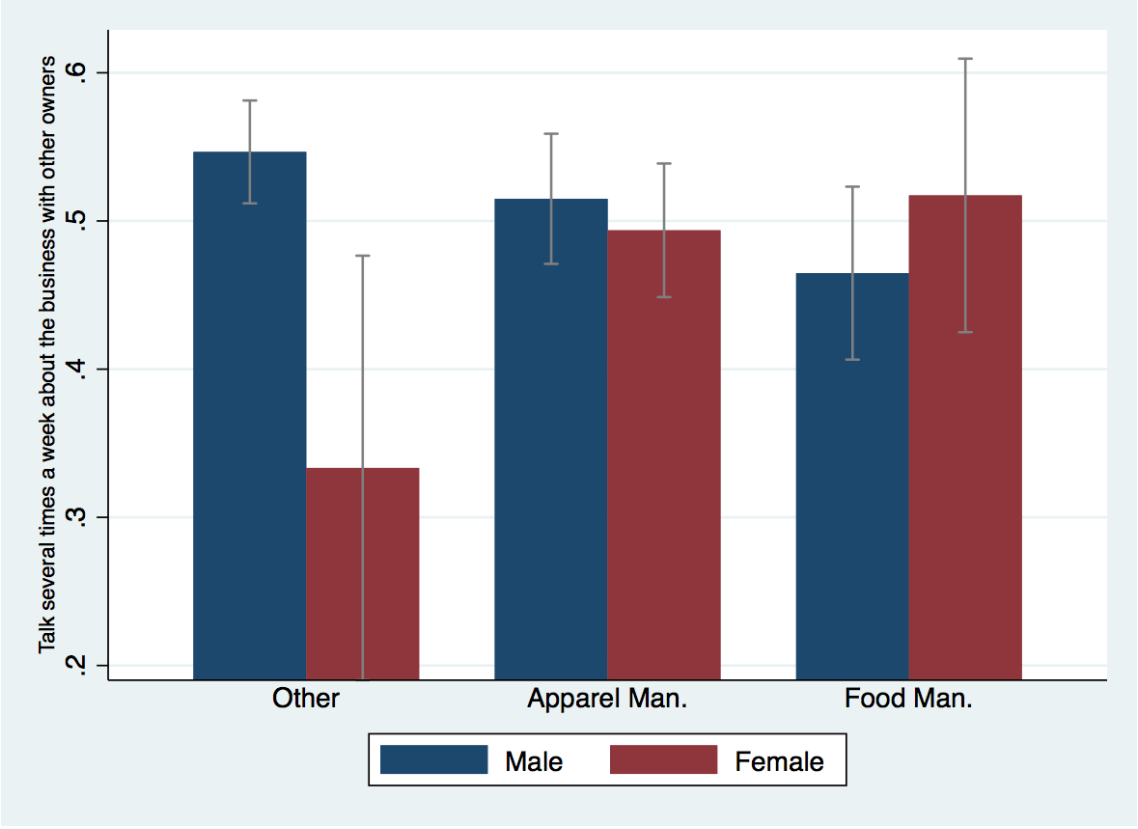
Source: 2016 Lusaka Census of Urban Entrepreneurs. N=1567



Source: 2016 Lusaka Census of Urban Entrepreneurs. N=629

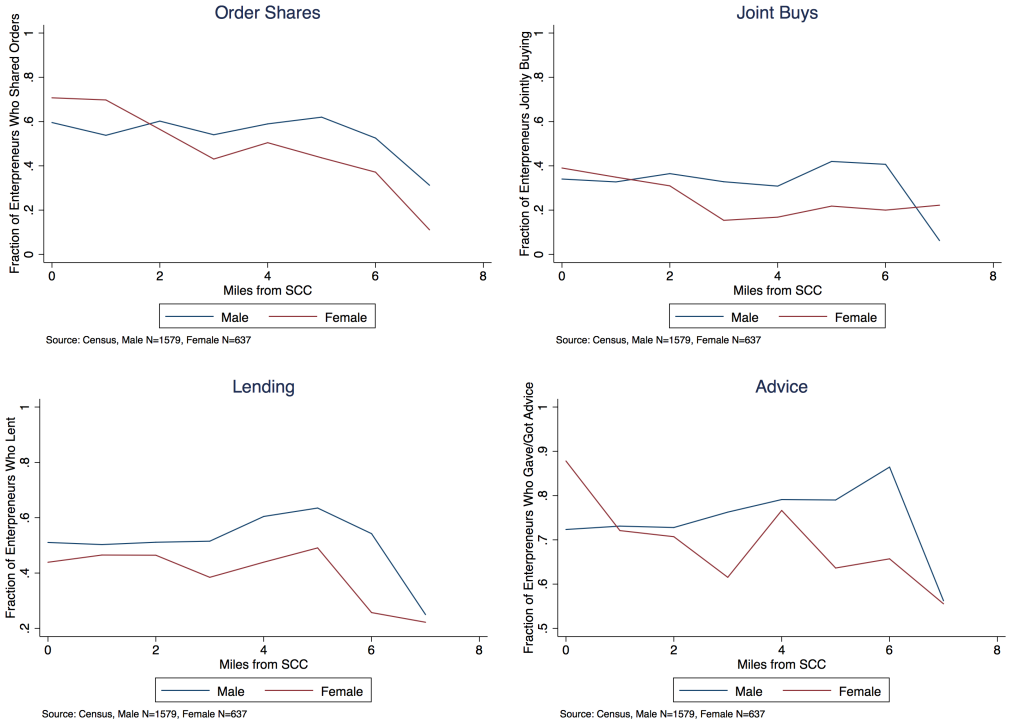
Note. This figure shows the different ways in which entrepreneurs learnt their job, by gender. Formal training was mainly interpreted as classroom training, usually delivered by governmental associations or NGOs. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

FIGURE 2.12: Talking about the business with other entrepreneurs



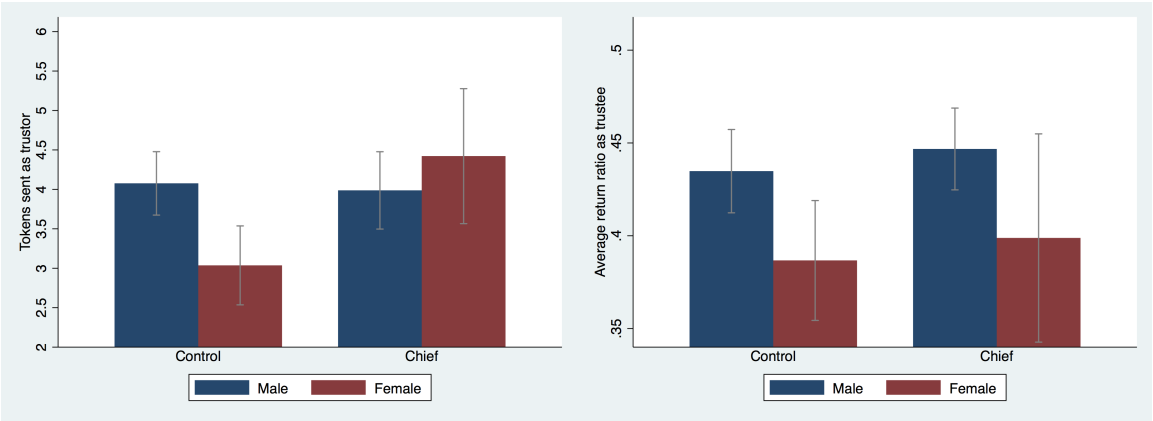
Note. This figure shows the proportion of men and women talking several times a week with other entrepreneurs about the business. We asked "Consider other business owners in your sector in this neighbourhood. How many times do you talk about topics related to the business?". The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

FIGURE 2.13: Cooperative activities by distance to SCC



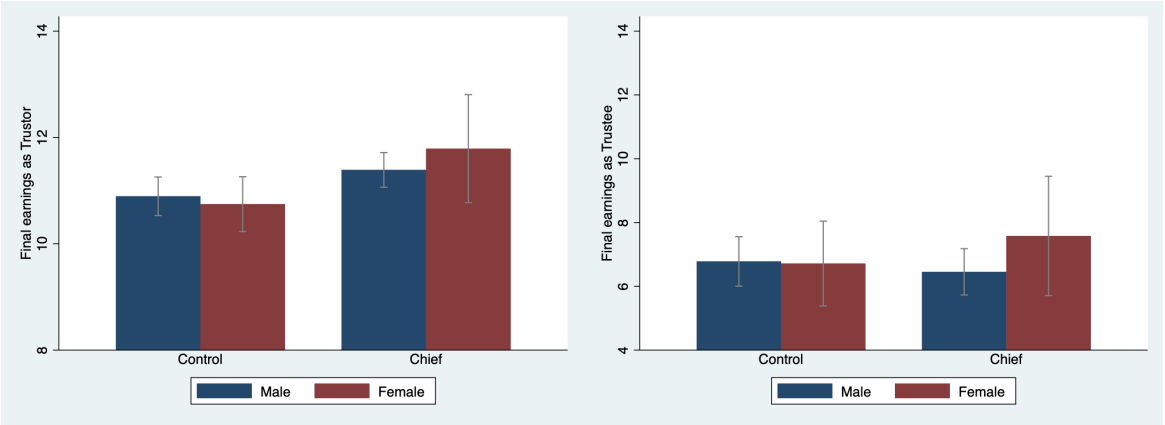
Note. This figure shows the fraction of entrepreneurs who said they cooperated with a similar business in the given activity by gender and distance from the SCC in 1-mile buckets. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

FIGURE 2.14: Number of tokens sent and average return ratio: control vs chief



Note. This figure shows the number of tokens sent (left) and the average return ratio (right) by gender and treatment group. The control group includes only businesses located inside markets. Data are from the experimental games.

FIGURE 2.15: Investor's and Trustee's final earnings: control vs chief



Note. This figure shows the final earnings of the Investor (left) and Trustee (right) by gender and institutional treatment group. The control group includes only businesses located inside markets. Data are from the experimental games.

2.10 Tables

TABLE 2.1: Countries by rule of law and discrimination in the family

Panel A: Number of countries			
		Discrimination in family	
		Low	*High*
Rule of law	*High*	47	26
		29.56%	16.35%
	Low	32	54
		20.13%	33.96%
Panel B: Female ownership			
		Discrimination in family	
		Low	*High*
Rule of Law	*High*	36%	18%
	sd	[0.13]	[0.12]
	N	13	11
	Low	24%	16%
	sd	[0.13]	[0.10]
	N	13	36

Note. Panel A of this table shows the proportion of firms with female ownership in each combination of countries, categorized using "discrimination of the family" from the SIGI above or below median (on the columns) and the WB rule of law rank decile (on the rows) above or below median. Panel B shows the categorization of countries in our data using "discrimination of the family" from the SIGI above or below median (on the columns) and the WB rule of law rank decile (on the rows) above or below median.

TABLE 2.2: Interaction between rule of law and female bargaining power

	Female ownership								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RoL$^{>med}$	0.017	0.033	0.027	0.020	0.017	-0.002	-0.146***	-0.152***	-0.183***
	(0.039)	(0.042)	(0.042)	(0.066)	(0.067)	(0.065)	(0.025)	(0.030)	(0.033)
Discr$^{<med}$	0.071*	0.086**	0.114**						
	(0.039)	(0.040)	(0.045)						
Discr$^{<med}$ x RoL$^{>med}$	0.113*	0.117*	0.115*						
	(0.063)	(0.063)	(0.064)						
Violence$^{<med}$				-0.050	-0.051	-0.047			
				(0.053)	(0.057)	(0.073)			
Violence$^{<med}$x RoL$^{>med}$				0.176*	0.165	0.161			
				(0.100)	(0.109)	(0.115)			
Gend Eq$^{>med}$							0.115**	0.106**	0.115**
							(0.051)	(0.053)	(0.050)
Gend Eq$^{>med}$ x RoL$^{>med}$							0.122*	0.110*	0.125**
							(0.062)	(0.060)	(0.055)
Log gdp pp (2011)		-0.015	-0.017		0.011	0.015		0.017	0.020
		(0.013)	(0.014)		(0.019)	(0.022)		(0.017)	(0.019)
Observations	73	73	73	36	36	36	52	52	52
R-squared	0.313	0.323	0.345	0.160	0.165	0.149	0.203	0.218	0.221
Sector	All	All	Manuf	All	All	Manuf	All	All	Manuf

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

Note. The regressions in this table test the theoretical prediction of complementarity between female bargaining power and fair and effective political institutions. Columns (1), (2) and (3) test the interaction between family discrimination (proxied by the SIGI index) and the World Bank rule of law rank percentile. The variable "RoL$^{>med}$" is equal to one for countries with rule of law above the sample median and the variable "Discr$^{<med}$" is equal to one for countries with discrimination in the family below the sample median. In regression (4), (5) and (6), we use the measure of violence against women from the WVS. The variable "Violence$^{<med}$" is equal to one for countries with violence against women below median. Regressions (2), (3) and (5), (6) and (8), (9) add controls for log of per capita income in 2011. Columns (7), (8) and (9) test whether rule of law is more important when combined with less bias against women, which we proxy using the overall WJP score on equal treatment of the genders by legal institutions. The variable "Gend Eq$^{>med}$" is equal to one for WJP score of gender equality above the sample median. Columns (3), (6) and (9) limit the sample of entrepreneurs in the WBES to manufacturing.

TABLE 2.3: Descriptive statistics for entire Census

	Mean	Observations
Number of Employees	2.27	40,517
Industry (Naics2)		
Retailing industry	0.51	48,163
Manufacturing industry	0.08	48,163
Accomodation/food services industry	0.14	48,163
Other Services	0.13	48,163
All Other Industries	0.13	48,163
Number of Employees by Industry		
Number Emp. Retail ind.	1.12	20,472
Number Emp. Manufacturing industry	3.86	3,625
Number Emp. accom/food services ind	2.34	5,854
Number Emp. other services ind	1.20	5,541
Number Emp. other ind	7.05	4,873
Business is part of a chain	0.04	48,670
Business has standalone structure (bricks)	0.11	48,670
Business in building with multiple businesses	0.71	48,670
Business has standalone structure (cardboard)	0.03	48,670
Business is in residential house	0.15	48,670

Note. This table presents industry composition for the 2016 Lusaka Census of Urban Entrepreneurs. All entries have been rounded to two decimal places. Observations may be dropped because of missing values.

TABLE 2.4: Education and firm size by gender

	Male Mean	Fem Mean	Diff	Male N	Fem N
Panel A: Firm Size					
No. full-time emp.	1.05	0.68	-0.38***	1579	635
	(0.06)	(0.07)	(0.10)		
No. part-time emp.	0.99	0.29	-0.69***	1571	633
	(0.08)	(0.04)	(0.13)		
No. apprentices/unpaid	0.68	0.35	-0.33	1575	634
	(0.13)	(0.04)	(0.21)		
No. family members	0.49	0.30	-0.19***	1546	602
	(0.03)	(0.03)	(0.04)		
Panel B: Education					
Training (mgmt/Entrep)	0.21	0.28	0.06***	1570	631
	(0.01)	(0.02)	(0.02)		
University	0.04	0.03	-0.01	1566	629
	(0.00)	(0.01)	(0.01)		
Diploma	0.16	0.19	0.03*	1566	629
	(0.01)	(0.02)	(0.02)		
Completed Secondary	0.23	0.20	-0.03	1566	629
	(0.01)	(0.02)	(0.02)		
No Formal Education	0.02	0.01	-0.01*	1566	629
	(0.00)	(0.00)	(0.01)		

Note. Panel A compares firm size by gender for completed intervews in the Manufacturers Survey. Panel B compares employment and education by gender for completed interviews in the Manufacturers Survey. All entries have been rounded to two decimal places. Sales variables contain outliers. Stars denote statistical significance of the two-sided t-test by gender. *** denotes $p<0.01$, ** denotes $p<0.05$, and * denotes $p<0.1$.

TABLE 2.5: Gender sales gap for all manufacturers

	Logged Sales							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Week		Week		Week		Week	
	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Female	-0.587***	-0.669***	-0.596***	-0.673***	-0.135**	-0.294***	-0.0832	-0.230***
	(0.0584)	(0.0709)	(0.0577)	(0.0712)	(0.0569)	(0.0698)	(0.0631)	(0.0765)
Apparel					-1.037***	-0.899***	-0.952***	-0.833***
					(0.0610)	(0.0742)	(0.0594)	(0.0746)
Food					0.304***	0.765***	0.195**	0.632***
					(0.0832)	(0.0928)	(0.0852)	(0.0956)
Hours Worked							0.0495***	0.0671***
							(0.0166)	(0.0198)
Days Worked							0.0599*	0.0827**
							(0.0312)	(0.0374)
Married							0.157***	0.135**
							(0.0568)	(0.0675)
Observations	2068	1848	2054	1836	2068	1848	2047	1830
Adjusted R^2	0.043	0.045	0.109	0.085	0.217	0.232	0.257	0.252
Mean Dep Var	7.1	5.89	7.09	5.89	7.1	5.89	7.1	5.88
Education Cnt	N	N	Y	Y	N	N	Y	Y
SD DV	1.28	1.44	1.28	1.43	1.28	1.44	1.27	1.43
F-test Educ = 0			0	0			0	0
P-val Food = Apparel					0	0	0	0

Robust SE in parentheses.* p<0.10, ** p<0.05, ***

Note. This table shows drivers of logged sales in a good or bad week for entrepreneurs of all industries in the 2016 Manufacturers Survey. The dependent variables are the log of the answers given when asked the sales in a good week and the sales in a bad week. If the person could not provide an exact number, we also asked for an upper and lower bound. We then imputed the average of the bounds to the sales variables. We regress on indicator variables indicating different levels of educational achievement: the omitted category is "illiterate or literate, no formal education", the other categories are "Primary Incomplete", "Primary", "Secondary Incomplete", "Secondary", "Diploma", "University". For industry, the omitted category is "other".

TABLE 2.6: Trust and cooperation in Zambia by gender

	Male Mean	Fem Mean	Diff	Male N	Fem N
Panel A: Trust					
Trust Strangers	0.29	0.20	-0.09***	1590	644
	(0.45)	(0.40)	(0.02)		
Trust Neighbors	0.54	0.43	-0.11***	1589	640
	(0.50)	(0.50)	(0.02)		
Trust GSS	1.10	1.04	-0.07***	1585	645
	(0.31)	(0.19)	(0.01)		
Feel safe with shop unattended	2.81	2.49	-0.32**	692	253
	(1.71)	(1.59)	(0.12)		
Panel B: Cooperation					
Joint Buy	0.35	0.27	-0.08***	1579	637
	(0.48)	(0.44)	(0.02)		
Lent	0.55	0.44	-0.11***	1579	637
	(0.50)	(0.50)	(0.02)		
Advice	0.76	0.71	-0.05**	1579	637
	(0.43)	(0.45)	(0.02)		
Share Order	0.58	0.54	-0.04*	1579	637
	(0.49)	(0.50)	(0.02)		
Coop Average	0.56	0.49	-0.07***	1579	637
	(0.33)	(0.34)	(0.02)		

Note. Panel A shows mean differences in trust between women and men, and Panel B shows mean differences in cooperation between women and men. The variables "Trust Strangers" and "Trust Neighbors" are measured on a scale from 1-4, and have been converted into dummy variables by combining low scores (1 and 2) and high scores (3 and 4). Low scores were given a value of 0, and high scores a value of 1. The "Feel safe leaving shop unattented" variable is measured in a scale from 1-5. Higher scores indicate that people disagree with the statement, and do not feel safe leaving their shop unattended. The variable "Trust GSS" is a dummy measuring whether most people can be trusted or not: a value of 1 indicates that "most people can be trusted", and a value of 0 indicates that "you cannot be too careful in dealing with people". The variables "Joint Buy", "Lent", "Advice" and "Share Order" are dummies that indicate whether a person ever engaged in the relevant activity. The variable "Coop Average" is an index of cooperative behaviour, calculated as a simple average of the four dummies "Joint Buy", "Lent", "Advice" and "Share Order". The sample for both panels stems from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.7: Learning history and sales

	Learnt from			Logged Sales		Employment
	(1) Owner	(2) Owner in same sector	(3) Family or friends	(4) Good Week	(5) Bad Week	(6) Full time
Female	-0.192***	-0.139***	-0.152***	-0.638***	-0.736***	-0.507***
	(0.0215)	(0.0192)	(0.0232)	(0.0943)	(0.114)	(0.166)
Informal learning				0.0601	0.0860	-0.0988
				(0.0788)	(0.0943)	(0.151)
Female×Informal learning				0.364***	0.343**	0.325
				(0.124)	(0.157)	(0.202)
Observations	2080	2080	2080	1961	1751	2078
Adjusted R^2	0.047	0.031	0.073	0.160	0.131	0.170
Mean Dep Var	.3	.22	.36	7.08	5.86	.91
SD Dep Var	.46	.41	.48	1.27	1.43	2.23
Business Density Controls	Y	Y	Y	Y	Y	Y
Owner Controls	Y	Y	Y	Y	Y	Y
Industry Fixed Effects	N	N	N	N	N	N

Robust SE in parentheses. * p<0.10, ** p<0.05, ***p<0.01

Note. The variables "Owner", "Owner in same sector", "Family or Friends" are dummies that indicate whether an entrepreneur learnt the job from each category. "Informal learning" is a dummy for whether the owner learnt the business from any of the previous three categories. Business Density Controls include a dummy variable for whether the business is located within 100 meters of a market, the total number of businesses within 100 meters, and the number of business from the same sector within 100 meters. Business Owner Controls includes owner's age, business age, how many days the business owner spends working in the business, educational dummies and whether business owner is married or not. In columns (4) and (5), the dependent variables are the log of the answers given when asked the sales in a good week and the sales in a bad week. If the person could not provide an exact number, we also asked for an upper and lower bound. We then imputed the average of the bounds to the sales variables. In Column (6), the dependent variable is the number of full-time employees. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.8: The market effect on cooperation

	(1) Share Order	(2) Lent	(3) Advice	(4) Joint Buy	(5) Coop Average
Female	-0.0956**	-0.156***	-0.0918**	-0.111***	-0.114***
	(0.0378)	(0.0379)	(0.0397)	(0.0381)	(0.0283)
In market	0.102***	0.131***	0.0437*	0.0563*	0.0832***
	(0.0269)	(0.0484)	(0.0252)	(0.0327)	(0.0283)
Female × In market	0.0922*	0.0750	0.0696*	0.0492	0.0715*
	(0.0507)	(0.0504)	(0.0377)	(0.0680)	(0.0401)
Observations	2216	2216	2216	2216	2216
Adjusted R^2	0.018	0.032	0.008	0.011	0.034
Business Density Controls	Y	Y	Y	Y	Y
Business Owner Controls	N	N	N	N	N
Industry FE	N	N	N	N	N
P-values					
In + Fem.×In = 0	0	0	.002	.05	0
Fem. + Fem. × In market = 0	.923	.036	.313	.164	.146

Clustered standard errors in parentheses. * p<0.10, ** p<0.05, ***p<0.01

Note. The variables "Joint Buy", "Lent", "Advice" and "Share" are dummies that indicate whether a person ever engaged in the relevant activity. The variable "Coop Average" is an index of cooperative behaviour, calculated as a simple average of the four dummies "Joint Buy", "Lent", "Advice" and "Share". Business Density Controls include a dummy variable for whether the business is located within 100 meters of a market, the total number of businesses within 100 meters, and the number of business from the same sector within 100 meters. Business Owner Controls includes whether the business owner keeps written business records of every purchase and sale made, whether her/his business is registered individually, whether the business owners trust their neighbours, how old the business is, how many days the business owner spends working in the business, age of business owner and whether business owner is married or not. Standard errors are clustered at the 1 squared-km area level. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.9: Effect of distance to SCC on cooperation

	(1) Share Order	(2) Lent	(3) Joint Buy	(4) Advice	(5) Coop Average
Female	0.104	-0.0574	0.0204	0.139***	0.0514
	(0.0635)	(0.0407)	(0.0714)	(0.0319)	(0.0348)
2-4 miles from SCC	-0.0115	-0.00324	-0.0115	-0.0354	-0.0154
	(0.0415)	(0.0344)	(0.0621)	(0.0342)	(0.0289)
Female × 2-4 miles from SCC	-0.0852	0.0227	-0.0532	-0.137***	-0.0633
	(0.0761)	(0.0488)	(0.0766)	(0.0386)	(0.0414)
>4 miles from SCC	-0.0236	0.0987**	0.000515	0.0298	0.0263
	(0.0356)	(0.0385)	(0.0630)	(0.0325)	(0.0285)
Female × >4 miles from SCC	-0.208***	-0.0896	-0.148*	-0.241***	-0.172***
	(0.0704)	(0.0586)	(0.0789)	(0.0414)	(0.0431)
Observations	2216	2216	2216	2216	2216
Adjusted R^2	0.007	0.014	0.007	0.009	0.015
Business Density Controls	Y	Y	Y	Y	Y
Business Owner Controls	N	N	N	N	N
Industry FE	N	N	N	N	N
P-values					
2-4mi + Fem. × 2-4mi = 0	.132	.582	.124	0	.004
>4mi + Fem. × >4mi = 0	0	.856	0	0	0
Fem. + Fem. × 2-4mi = 0	.612	.409	.148	.969	.664
Fem. + Fem. × >4mi = 0	.001	.001	0	0	0

Clustered standard errors in parentheses. * p<0.10, ** p<0.05, ***p<0.01

Note. The variables "Joint Buy", "Lent", "Advice" and "Share" are dummies that indicate whether a person ever engaged in the relevant activity. The variable "Coop Average" is an index of cooperative behaviour, calculated as a simple average of the four dummies "Joint Buy", "Lent", "Advice" and "Share". Business Density Controls include a dummy variable for whether the business is located within 100 meters of a market, the total number of businesses within 100 meters, and the number of business from the same sector within 100 meters. Business Owner Controls includes whether the business owner keeps written business records of every purchase and sale made, whether her/his business is registered individually, whether the business owners trust their neighbours, how old the business is, how many days the business owner spends working in the business, age of business owner and whether business owner is married or not. Standard errors are clustered at the 1 squared-km area level. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.10: Effect of being in market on sales

	(1)	(2)	(3)
		Log Sales	
	Good Week	Bad Week	Yesterday
Female	-0.282***	-0.542***	-0.161
	(0.0943)	(0.115)	(0.111)
In market	-0.114	-0.113	-0.235***
	(0.0707)	(0.0721)	(0.0874)
Female × In market	0.200**	0.335***	0.245*
	(0.0917)	(0.117)	(0.134)
Observations	1911	1691	1364
Adjusted R^2	0.335	0.297	0.326
Business Density Controls	Y	Y	Y
Business Owner Controls	Y	Y	Y
Industry FE	Y	Y	Y
P-values			
In + Fem.×In = 0	.271	.062	.925
Fem. + Fem. × In market = 0	.233	.013	.345

Clustered standard errors in parentheses. * p<0.10, ** p<0.05, ***p<0.01

Note. The dependent variables are the log of the answers given when asked the sales in the previous working day, the sales in a good week and the sales in a bad week. If the person could not provide an exact number, we also asked for an upper and lower bound. We then imputed the average of the bounds to the sales variables. Business Density Controls include a dummy variable for whether the business is located within 100 meters of a market, the total number of businesses within 100 meters, and the number of business from the same sector within 100 meters. Business Owner Controls includes whether the business owner keeps written business records of every purchase and sale made, whether her/his business is registered individually, whether the business owners trust their neighbours, how old the business is, how many days the business owner spends working in the business, age of business owner and whether business owner is married or not. The regressions also include industry fixed-effects, a dummy on the owner's gender, a dummy on whether the business is located inside a market, and the interaction between the two of them. Standard errors are clustered at the 1 squared-km area level. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.11: Effect of distance to SCC on sales

	(1)	(2)	(3)
		Log Sales	
	Good Week	Bad Week	Yesterday
Female	-0.288***	-0.480***	0.102
	(0.0550)	(0.0914)	(0.186)
2-4 miles from SCC	-0.0558	-0.0420	-0.00167
	(0.157)	(0.156)	(0.0936)
Female × 2-4 miles from SCC	0.179*	0.286**	-0.0380
	(0.0942)	(0.117)	(0.198)
>4 miles from SCC	-0.253	-0.141	-0.196**
	(0.161)	(0.164)	(0.0892)
Female × >4 miles from SCC	0.125	0.0876	-0.141
	(0.0864)	(0.113)	(0.197)
Observations	1911	1691	1364
Adjusted R^2	0.338	0.297	0.327
Business Density Controls	Y	Y	Y
Business Owner Controls	Y	Y	Y
Industry FE	Y	Y	Y
P-values			
2-4mi + Fem. × 2-4mi = 0	.37	.269	.8
>4mi + Fem. × >4mi = 0	.314	.817	.037
Fem. + Fem. × 2-4mi = 0	.252	.017	.52
Fem. + Fem. × >4mi = 0	.025	0	.698

Clustered standard errors in parentheses. * p<0.10, ** p<0.05, ***p<0.01

Note. The dependent variables are the log of the answers given when asked the sales in the previous working day, the sales in a good week and the sales in a bad week. If the person could not provide an exact number, we also asked for an upper and lower bound. We then imputed the average of the bounds to the sales variables. Business Density Controls include a dummy variable for whether the business is located within 100 meters of a market, the total number of businesses within 100 meters, and the number of business from the same sector within 100 meters. Business Owner Controls includes whether the business owner keeps written business records of every purchase and sale made, whether her/his business is registered individually, whether the business owners trust their neighbours, how old the business is, how many days the business owner spends working in the business, age of business owner and whether business owner is married or not. The regressions also include industry fixed-effects, a dummy on the owner's gender, dummies on the business' distance to the Small Claims Court, and their interaction. Standard errors are clustered at the 1 squared-km area level. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.12: Experimental treatments

	Institutional Access Treatment		
	Control	Chief	SCC
Inside/Inside	95	96	96
Outside/Inside	94	0	96

Note. Numbers indicate the number of rounds in each treatment cell. Each player played two rounds (once as Investor and once as Trustee), with two distinct entrepreneurs. Three out of the 480 rounds that occurred were not used for data quality reasons.

TABLE 2.13: Games' behaviour and earnings: control vs chief

	(1) Tokens Investor Sent	(2) Average Return Ratio	(3) Investor's Earnings	(4) Trustee's Earnings
Female	-1.243***	0.000115	-0.210	0.355
	(0.430)	(0.0263)	(0.400)	(1.202)
Chief	-0.0851	0.0247	0.505	-0.373
	(0.382)	(0.0200)	(0.317)	(0.668)
Female × Chief	1.610**	-0.0524	0.743	1.055
	(0.682)	(0.0417)	(0.694)	(1.552)
Observations	189	190	189	190
Adjusted R^2	0.020	0.000	0.029	-0.000
Mean	4.09	.42	10.86	6.76
SD	2.412	.135	2.244	4.519

Robust standard errors in parentheses. * p<0.10, ** p<0.05, ***p<0.01

Note. The dependent variable in Column (1) is the number of tokens sent by the Investor and in Column (2) is the average return ratio by the Trustee. The dependent variables in Columns (3) and (4) are the Investor's and Trustee's earnings, respectively. All regressions include a dummy for the round played and exclude the control group in the "outsider" condition and the Small Claims Court treatment. The sample comes from our lab-in-the-field games conducted in Lusaka in 2017.

2.A Proofs

Proof of Proposition 1:

A male P will always shirk when shirking is disputable. When shirking is indisputable, he will not shirk as long as $s\pi > b$. We refer to this as the IC contract and it must be satisfied for a contract to occur, since the woman will never partner when there is always shirking.

If the IC constraint is satisfied, then in any contract between a female E and a male P, the man will in expectation receive in expectation $s\pi - q + \delta\, b$, and so s must also satisfy $s > \frac{q-\delta b}{\pi}$. The woman will receive $(1-s)\pi - q - \delta(b+\Delta)$, and so her participation requires that $s < 1 - \frac{q+\delta(b+\Delta)}{\pi}$.

The total expected surplus is $\pi - 2q - \delta\Delta$, and that must be positive (or $\delta < \frac{\pi-q}{\Delta}$) for a partnership to occur. We refer to this as the individual rationality of IR constraint.

If the IC constraint is slack, then men and women share the surplus according to the bargaining rule so that men receive $\beta(\pi - 2q - \delta\Delta)$ and women receive $(1-\beta)(\pi - 2q - \delta\Delta)$, which implies that $s = \beta - \frac{(2\beta-1)q+\beta\delta\Delta+\delta b}{\pi}$, which satisfies $1 - \frac{q+\delta(b+\Delta)}{\pi} > s > \frac{q-\delta b}{\pi}$as long as $\delta < \frac{\pi-q}{\Delta}$.

This value of s will satisfy the IC constraint if and only if $\beta\pi - (2\beta-1)\, q - \beta\delta\Delta - \delta\, b > b$ or $\frac{\beta(\pi-2q)+q-b}{\beta\Delta+b} > \delta$. If that condition fails to hold, then E must effectively pay P efficiency wages to eliminate shirking in the indisputable case. In that case, $s = \frac{b}{\pi}$, and the male receives $(1+\delta)b - q > \beta(\pi - 2q - \delta\Delta)$ and the woman receives $\pi - b - q - \delta(b+\Delta) < (1-\beta)(\pi - 2q - \delta\Delta)$. When $\pi < q + (1+\delta)\, b + \delta\Delta$ or $\frac{\pi-q-b}{\Delta+b} < \delta$, then the woman earns zero profits and the partnership will not occur.

If $\pi - 2q < \Delta\frac{q-b}{b}$ then $\frac{\pi-2q}{\Delta} < \frac{\beta\pi-(2\beta-1)q-b}{\beta\Delta+b}$ and if the contract satisfies the IR constraint it automatically satisfies the IC constraint as well. Consequently, if $\frac{\pi-2q}{\Delta} > \delta$ the contract specifies $s = \beta - \frac{(2\beta-1)q+\beta\delta\Delta+\delta b}{\pi}$, giving expected welfare of $\beta(\pi - 2q - \delta\Delta)$ to P and $(1-\beta)(\pi - 2q - \delta\Delta)$ to E. If $\frac{\pi-2q}{\Delta} < \delta$ then there is no contract.

If $\pi - 2q > \Delta\frac{q-b}{b}$, which will always hold when $q < b$, then we have$\frac{\pi-2q}{\Delta} > \frac{\pi-q-b}{\Delta+b} > \frac{\beta(\pi-2q)+q-b}{\beta\Delta+b}$. If $\frac{\beta(\pi-2q)+q-b}{\beta\Delta+b} > \delta$, then the contract specifies $s = \beta - \frac{(2\beta-1)q+\beta\delta\Delta+\delta b}{\pi}$, giving expected welfare of $\beta(\pi - 2q - \delta\Delta)$ to P and $(1-\beta)(\pi - 2q - \delta\Delta)$ to E. If $\frac{\pi-q-b}{\Delta+b} > \delta > \frac{\beta(\pi-2q)+q-b}{\beta\Delta+b}$, then the contract specifies $s = \frac{\pi}{b}$ and the expected male payout is $(1+\delta)\, b - q$ and the expected female payout is $\pi - b(1+\delta) - q - \delta\Delta$. If $\frac{\pi-q-b}{\Delta+b} < \delta$, then there is no contract.

Proof of Proposition 2:

Women enter if and only if expected returns, denoted R, are greater than $\theta(.5\pi - q)$. There are three cases to consider. If $\frac{\pi-2q}{\Delta} < \frac{q}{b} - 1$, and $\frac{\pi-2q}{\Delta} < \delta$, or $\frac{\pi-2q}{\Delta} > \frac{q}{b} - 1$ and $\frac{\pi-q-b}{\Delta+b} < \delta$ then women do not partner with men. As $\frac{\pi-q-b}{\Delta+b} > \frac{\pi-2q}{\Delta}$ if and only if $\frac{\pi-2q}{\Delta} > \frac{q}{b} - 1$, these conditions are satisfied whenever $\delta > Max\left[\frac{\pi-2q}{\Delta}, \frac{\pi-q-b}{\Delta+b}\right]$. In that case, women enter if and only if $1 - \theta > m_i$.

If $\frac{\pi-2q}{\Delta} < \frac{q}{b} - 1$, and $\frac{\pi-2q}{\Delta} > \delta$, or $\frac{\pi-2q}{\Delta} > \frac{q}{b} - 1$ and $\frac{\beta(\pi-2q)+q-b}{\beta\Delta+b} > \delta$, then women receive $(1-\beta)(\pi-2q-\delta\Delta)$ when they partner with a men. As $\frac{\beta(\pi-2q)+q-b}{\beta\Delta+b} < \frac{\pi-2q}{\Delta}$ if and only if $\frac{\pi-2q}{\Delta} > \frac{q}{b} - 1$, these conditions are satisfied whenever $\delta < Min\left[\frac{\pi-2q}{\Delta}, \frac{\beta(\pi-2q)+q-b}{\beta\Delta+b}\right]$. In that case female returns from entry equal $(0.5\pi - q)(1-m_i) + m_i(1-\beta)(\pi-2q-\delta\Delta)$. If $1 - \frac{\theta}{2(1-\beta)} > \frac{\delta\Delta}{\pi-2q}$, then women enter for all values of m_i. If $\left(1 - \frac{\theta}{2(1-\beta)}\right)(\pi-2q) < \delta\Delta$, then women enter if and only if $m_i < m^* = \frac{(\pi-2q)(1-\theta)}{(\pi-2q)(2\beta-1)+2(1-\beta)\delta\Delta}$. The value of m^* is rising with π, falling with q, θ, δ, Δ and β.

If $\frac{\pi-2q}{\Delta} > \frac{q}{b} - 1$ and $\frac{\pi-q-b}{\Delta+b} > \delta > \frac{\beta(\pi-2q)+q-b}{\beta\Delta+b}$, then the expected female payout from partnering with a man is $\pi - b(1+\delta) - q - \delta\Delta$, and women enter if and only if

$(0.5\pi - q)(1-m_i) + m_i(\pi - b(1+\delta) - q - \delta\Delta) > \theta(0.5\pi - q)$. In that case, women always enter if $1 - 0.5\theta > \frac{b(1+\delta)+\delta\Delta}{\pi-2q}$. If $(1-0.5\theta)(\pi-2q) < b(1+\delta) + \delta\Delta$, then women enter if and only if $m_i < m^{**} < \frac{(\pi-2q)(1-\theta)}{2b(1+\delta)+2\delta\Delta-\pi}$. The value of m^{**} is rising with π, falling with q, b, θ, δ, and Δ.

2.B Games procedures

Six surveyors and two recruiters were hired to conduct the games and were managed by a research assistant. Typically two days were spent in each market. We created two lists of randomly-ordered businesses for each market; one list for businesses located inside the market and one for businesses located outside the market. Manufacturers, being of primary interest, were placed at the top of each list. The two recruiters met early and were given the randomized list of entrepreneurs; they then set up appointments with potential participants, following the order of the list. If a marketeer could not be located or refused to participate, the following skip pattern was implemented: the marketeer was replaced by the next female on the list and as soon as all women had been exhausted, the recruiters simply moved to the next marketeer on the list. In markets where the response rate was low, or we failed to find many businesses or an above average amount of businesses were closed, we thus also surveyed and played the games with non-manufacturers. It is also important to note that all entrepreneurs on the outside list were screened extensively to ensure that they truly did not belong to the market.

In general, the first day at a market, the games were played with inside-inside pairs, wherein both players worked within the market. The corresponding three treatments arms for these pairs were control, SCC and chief. Generally, the second day at a market, the games were played with inside-outside pairs, wherein one player was a marketeer and the other an entrepreneur located outside of the market. The corresponding two treatment arms for these pairs were control and SCC. Surveys were conducted first, and then the games. However, if there were delays in reaching a participant or if a player refused to play the games after having completed the survey, the order was switched. The order in which each treatment arm was conducted was randomized in each market

to limit selection bias due to logistical factors (E.g. Time of day, eagerness of participants, etc.).

We ran the games on groups of six entrepreneurs at time, with one surveyor assigned to one entrepreneur. For each market, a switch matrix was constructed to inform surveyors whether their assigned entrepreneur was to be Player A (Investor) or Player B (Trustee) first, and who the entrepreneur was to play in each round. For logistical reasons, surveyors whose respondents were located outside of the market always started the games with the respondent being the Trustee, therefore all outsiders played the games in the same order, first as Trustees then as Investors. Once finishing a survey with a respondent, a surveyor sent a text to their corresponding surveyor to inform that they were about to start the instructions for the games in the following format: "surveyornum-initials-CensusID-AB/BA-start (e.g. 4-D-6230-AB-start)". The instructions for the games were given to the respondent in written as well as oral form in the respondent's preferred language, and surveyors asked the respondent check questions to ensure that they fully understood the pay-out rules, who they were playing against, and the possibilities for complaint when relevant. Once both surveyors in a pair had received the "ready" text message from their partner surveyor for that round and had ensured that the participant understood the rules, they started with the game.

The surveyor of the Investor sent the surveyor of the corresponding Trustee the number of tokens his player had chosen to send in the following format "SurveyorNum-Surname-Game-Player-TokensSent (e.g. 4-AD-R1-A3)". The surveyor of the Trustee then responded with the number of tokens the Trustee had decided to send back in the following format "SurveyorNum-Surname-Game-Player-TokensReturned (e.g. 2-PB-R1-B2)". The surveyors were instructed to always use neutral language to inform the respondents of the amount that had been sent (or sent back) to them.

In the case of the SCC and chief treatments, the Investor, upon being told how many tokens the Trustee had sent back, was asked whether they wanted to complain to the SCC (chief) or not. The surveyor then messaged the corresponding surveyor whether or not the Investor intended to complain (format: A-Comp OR A-NoComp). If the Investor complained, the surveyor also messaged the recruiter in the following format: "SurveyorNum-Surname-ANumGiven-BNumReturned-TotNum-Co E.g. 4-AD-A3-B2-Tot9-Comp". In the case of the SCC treatment, the recruitment officer already had a completed matrix of decisions from the SCC (obtained from the SCC before the start of the games field work). The recruitment officer examined the matrix and sent the SCC's decision to the surveyor of the Investor in the following format: "ANumTokensReceived-BNumTokensReceived. E.g. A6-B3". In the case of the chief treatment, the recruiter would ask the market's chief in real-time how s/he wanted to settle the complaint and sent their response in the same format to the surveyor of the Investor. The chief was given information (both oral and written) prior to the games commencing, which explained the

game and made clear her/his role as an arbitrator during the games. Similar to the respondents, they were asked questions to check whether they had fully understood the games and their role in the games. If/ when a complaint reached the chief, the recruiter who was assigned to her/him would announce the complaint and ask for her/his decision in the following way: "There has been a complaint from a player. In this game, player A (the Investor) sent XX number of tokens to B (the Trustee), which means B received XX number of tokens. B sent back XX tokens. A has complained to you. Do you wish to redistribute the tokens? If so, how?"

It is important to note, that the players were never told who they were playing against, however, they knew whether the player was located inside or outside the market. Furthermore, after playing the first round as Investor or Trustee, they played the second round as Trustee (Investor respectively) against a new player, so that no two players played each other twice. It was explicitly made clear to them that they would be randomly assigned to a new partner after the first round and they were reminded whether this player was inside or outside the market and if they (or the opposing player) could complain to the chief or SCC or neither.

As some of the markets were quite small, and six surveyors, two recruiters and one research assistant naturally stuck out, it is possible that participants had heard of the games we were conducting before we reached them. Thus, we cannot fully exclude the possibility that some players may have played with more information to begin with i.e. on how the SCC or chief had decided in other cases in the market, though, this is more likely to be true for the second day in a specific market as the games conducted in a specific day followed back-to-back, thus there was not much time for entrepreneurs to gossip in-between. Furthermore, the field team sought to minimize the attention drawn to themselves by merely sending the two recruiters into the depths of the markets to find participants, whilst the surveyors would generally just directly be taken from one entrepreneur to the other. Data consistency checks were conducted at the end of each field day.

2.C Additional tables and robustness checks

TABLE 2.C.1: Cross-country correlations with female ownership

Variables	Correlation coefficient	T-stat univar reg	T-stat bivar reg	N
Panel A - discrimination, norms and rule of law				
Discrimination in the family (SIGI-19)	-0.62	-7.7	-8.01	75
Restricted physical integrity (SIGI-19)	-0.5	-4.41	-3.87	64
Men better in business than women (WVS04-14)	-0.64	-5.16	-3.98	36
Justifiable for men to beat wives (WVS04-14)	-0.29	-1.69	-1.22	36
Health and safety rank (GGG-16)	-0.48	-4.74	-4.37	69
Rule of law percentile rank (WB WGI-17)	0.32	3.28	1.94	82
Rule of Law Score (WJP-17)	0.27	2.21	0.67	50
Female disadvantage with police (GPP-17)	-0.31	-2.05	-1.26	53
Gender equality in courts (WJP experts)	0.35	2.93	1.9	52
Gender equality score (WJP experts)	0.3	2.29	1.19	52
Panel B - democracy, religion and education				
Average Polity and Democracy Index (1960-2000)	0.26	2.59	0.86	83
Democracy Index (EIU-18)	0.35	3.67	2.31	76
Female school enrollment, tertiary (UN-15)	0.25	2.16	0.79	56
Female school enrollment, secondary (UN-15)	0.52	5.94	4.02	56
Educational Attainment rank (GGG16)	-0.5	-5.86	-5.09	69
Percentage of Muslim citizens (UN-00)	-0.69	-10.68	-10.39	84
Panel C - ease of doing business				
Ease of doing business score global (DB10-14)	0.29	2.96	1.2	56
Score-Enforcing contracts (DB04-15)	0.38	3.44	2.74	61
Score-Registering property (DB05-15)	0.33	2.89	2.18	61

Note. This Table shows the correlation between empirical measures of the model parameters and female ownership across countries. Column 1 reports the raw correlation coefficient. Column 2 reports the t-stat of a univariate regression of female entrepreneurship on the given row variable. Column 3 reports the t-stat of a bivariate regression of female entrepreneurship on the given row variable and log GDP per capita in 2011. The last column reports the number of observations, which may vary depending on the surveys' coverage and years. Acronym in brackets show the source of the variable and the years of collection used.

TABLE 2.C.2: Correlations between trust and cooperation

	Trust ST	Trust NB	Trust GSS	Joint Buy	Lent	Advice	Share Order	Coop Avg
Trust Stranger	1							
Trust Neighbor	0.313***	1						
Trust GSS	0.129***	0.119***	1					
Joint Buy	0.0516*	0.0577**	0.00748	1				
Lent	0.0773***	0.123***	0.0151	0.259***	1			
Advice	0.0596**	0.0978***	-0.0123	0.258***	0.356***	1		
Share Order	0.0618**	0.0819***	-0.0394	0.349***	0.328***	0.347***	1	
Coop Average	0.0901***	0.129***	-0.0104	0.665***	0.708***	0.683***	0.734***	1
Complexity	0.0214	-0.00476	0.0435*	0.103***	0.0685**	0.0209	-0.0330	0.0569**
N	2216							

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. This table shows correlations between the trust variables and cooperation variables. The variables "Trust Strangers" and "Trust Neighbors" are measured on a scale from 1-4, and have been converted into dummy variables by combining low scores (1 and 2) and high scores (3 and 4). Low scores were given a value of 0, and high scores a value of 1. The variable "Trust GSS" is a dummy measuring whether most people can be trusted or not: a value of 1 indicates that "most people can be trusted", and a value of 0 indicates that "you cannot be too careful in dealing with people". The variables "Joint Buy", "Lent", "Advice" and "Share" are dummies that indicate whether a person ever engaged in the relevant activity. The variable "Coop Average" is an index of cooperative behaviour, calculated as a simple average of the four dummies "Joint Buy", "Lent", "Advice" and "Share". The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.C.3: Correlations between sales and trust or cooperative behaviour

	Logged sales		
	Good Week	Bad Week	Yesterday
Complexity	.3927***	.3806***	.4030***
Trust NB	.0128	-.0027	-.0201
Trust ST	.0415*	.0376	.0459*
Trust GSS	.0405	.0439*	.0140
Coop Average	.0840***	.0189	.0755***
Lent	.0286	-.0275	.0224
Advice	.0508**	.0110	.0441*
Share Order	.0587***	-.0007	.0315
Joint Buy	.0973***	.0734***	.1178***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. This table shows correlations between sales and trust or cooperative behaviour. The dependent variables are the log of the answers given when asked the sales in the previous working day, the sales in a good week and the sales in a bad week. If the person could not provide an exact number, we also asked for an upper and lower bound. We then imputed the average of the bounds to the sales variables. The variables "Trust Strangers" and "Trust Neighbors" are measured on a scale from 1-4, and have been converted into dummy variables by combining low scores (1 and 2) and high scores (3 and 4). Low scores were given a value of 0, and high scores a value of 1. The variable "Trust GSS" is a dummy measuring whether most people can be trusted or not: a value of 1 indicates that "most people can be trusted", and a value of 0 indicates that "you cannot be too careful in dealing with people". The variables "Joint Buy", "Lent", "Advice" and "Share" are dummies that indicate whether a person ever engaged in the relevant activity. The variable "Coop Average" is an index of cooperative behaviour, calculated as a simple average of the four dummies "Joint Buy", "Lent", "Advice" and "Share". The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.C.4: The market effect on cooperation with industry fixed-effects

	(1) Share Order	(2) Lent	(3) Advice	(4) Joint Buy	(5) Coop Average
Female	-0.0453	-0.0314	-0.0488	-0.0367	-0.0406
	(0.0494)	(0.0426)	(0.0481)	(0.0463)	(0.0369)
In market	0.0569*	0.104***	0.0351	0.0820**	0.0694***
	(0.0311)	(0.0363)	(0.0281)	(0.0311)	(0.0250)
Female×In market	0.0877	0.0905*	0.0895**	0.0538	0.0804*
	(0.0569)	(0.0456)	(0.0401)	(0.0659)	(0.0421)
Observations	2077	2077	2077	2077	2077
Adjusted R^2	0.059	0.090	0.028	0.037	0.088
Business Density Controls	Y	Y	Y	Y	Y
Business Owner Controls	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
		P-values			
In+Fem.× In=0	.007	0	.001	.034	0
Fem. + Fem. × In market = 0	.232	.055	.122	.612	.082

Clustered standard errors in parentheses. p<0.10, ** p<0.05, *** p<0.01

Note. The dependent variables are our measures of cooperation. The variables "Joint Buy", "Lent", "Advice" and "Share" are dummies that indicate whether a person in our Manufacturers Survey sample ever engaged in the relevant activity. The variable "Coop Average" is an index of cooperative behaviour, calculated as a simple average of the four dummies "Joint Buy", "Lent", "Advice" and "Share". Business Density Controls include a dummy variable for whether the business is located within 100 meters of a market, the total number of businesses within 100 meters, and the number of business from the same sector within 100 meters. Business Owner Controls includes whether the business owner keeps written business records of every purchase and sale made, whether her/his business is registered individually, whether the business owners trust their neighbours, how old the business is, how many days the business owner spends working in the business, age of business owner and whether business owner is married or not. The regressions also include industry fixed-effects, a dummy on the owner's gender, a dummy on whether the business is located inside a market, and the interaction between the two of them. Standard errors are clustered at the 1 squared-km area level. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.C.5: Effect of distance to SCC on cooperation with industry fixed-effects

	(1) Share Order	(2) Lent	(3) Joint Buy	(4) Advice	(5) Coop Average
Female	0.188***	0.0218	0.120***	0.186***	0.129***
	(0.0485)	(0.0471)	(0.0433)	(0.0258)	(0.0287)
2-4 miles from SCC	0.0298	0.0172	0.0136	-0.0311	0.00739
	(0.0396)	(0.0383)	(0.0510)	(0.0399)	(0.0351)
Female × 2-4 miles from SCC	-0.144**	0.000277	-0.0938*	-0.146***	-0.0960**
	(0.0583)	(0.0621)	(0.0512)	(0.0419)	(0.0381)
>4 miles from SCC	0.00991	0.0515	-0.0372	0.0333	0.0144
	(0.0335)	(0.0343)	(0.0423)	(0.0363)	(0.0292)
Female × >4 miles from SCC	-0.230***	-0.00737	-0.165***	-0.233***	-0.159***
	(0.0521)	(0.0644)	(0.0546)	(0.0338)	(0.0356)
Observations	2077	2077	2077	2077	2077
Adjusted R^2	0.057	0.076	0.035	0.028	0.074
Business Density Controls	Y	Y	Y	Y	Y
Business Owner Controls	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y
		P-value			
2-4mi + Fem. × 2-4mi = 0	.014	.642	.031	0	.001
>4mi + Fem. × >4mi = 0	0	.355	0	0	0
Fem. + Fem. × 2-4mi = 0	.345	.585	.231	.337	.253
Fem. + Fem. × >4mi = 0	.148	.763	.196	.186	.269

Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Note. The dependent variables are our measures of cooperation. The variables "Joint Buy", "Lent", "Advice" and "Share" are dummies that indicate whether a person in our Manufacturers Survey sample ever engaged in the relevant activity. The variable "Coop Average" is an index of cooperative behaviour, calculated as a simple average of the four dummies "Joint Buy", "Lent", "Advice" and "Share". Business Density Controls include a dummy variable for whether the business is located within 100 meters of a market, the total number of businesses within 100 meters, and the number of business from the same sector within 100 meters. Business Owner Controls includes whether the business owner keeps written business records of every purchase and sale made, whether her/his business is registered individually, whether the business owners trust their neighbours, how old the business is, how many days the business owner spends working in the business, age of business owner and whether business owner is married or not. The regressions also include industry fixed-effects, a dummy on the owner's gender, dummies on the business' distance to the Small Claims Court, and their interaction. Standard errors are clustered at the 1 squared-km area level. The sample comes from our Manufacturers Survey in Lusaka (N=2,216).

TABLE 2.C.6: Women and help from institutions

	Male Mean	Fem Mean	Diff	Male N	Fem N
Panel A: Knowledges and Opinions					
Heard of SCC	0.22	0.20	-0.02	347	128
	(0.41)	(0.40)	(0.04)		
Would Leave Shop	0.44	0.32	-0.12**	347	127
	(0.50)	(0.47)	(0.05)		
Opposition Abused	0.37	0.41	0.04	339	125
	(0.48)	(0.49)	(0.05)		
Panel B: Work Disputes					
Any Dispute	0.67	0.60	-0.07	348	129
	(0.47)	(0.49)	(0.05)		
Disp. Workplace	0.19	0.12	-0.08*	348	129
	(0.39)	(0.32)	(0.04)		
Disp. Bus. Agree	0.32	0.26	-0.06	348	128
	(0.47)	(0.44)	(0.05)		
Disp. Over Debt	0.52	0.40	-0.12**	347	129
	(0.50)	(0.49)	(0.05)		
Disp. Over Goods	0.34	0.25	-0.09*	347	129
	(0.47)	(0.43)	(0.05)		
Panel C: Theft, Assault, Harassment					
Victim of Theft	0.40	0.37	-0.03	348	129
	(0.49)	(0.49)	(0.05)		
Thief Caught	0.22	0.06	-0.16**	140	48
	(0.42)	(0.24)	(0.06)		
Victim of Assault	0.05	0.03	-0.02	348	129
	(0.22)	(0.17)	(0.02)		
Assailant Caught	0.22	0.00	-0.22	18	3
	(0.43)	(0.00)	(0.25)		
Police Harass	0.05	0.00	-0.05**	348	129
	(0.22)	(0.00)	(0.02)		

Note. This table shows mean differences in the access and use of institutions between women and men. All the variables are dummy variables. The variable "Heard of SCC" is 1 if the person has heard of the SCC, and 0 otherwise. The variable "Would Leave Shop" is 1 if the person would feel comfortable leaving the shop unattended for 30 minutes, and 0 otherwise. The variable "Opposition Abused" is 1 if the person feels that members of the (political) opposition frequently receive verbal or physical abuse, and 0 otherwise. The work dispute variables are equal to 1 if the person has experienced the mentioned type of work-disputed, and 0 otherwise. The variables "Victim of theft" and "Victim of assault" equal 1 if the person has been a victim of these crimes, and 0 otherwise. The variables "Thief caught" and "Perpetrator caught" equal 1 if the crime perpetrator has been captured, and 0 otherwise. The variable "Police Harass" is 1 if the person has suffered from police harassment in the last 12 months, and 0 otherwise. The sample stems from the survey we conducted with business owners before the lab-in-the-field games.

TABLE 2.C.7: Tokens investor sends by institutional treatment

	Tokens Investor sent		
	(1)	(2)	(3)
Female	-1.243***	-1.030***	-1.236***
	(0.430)	(0.331)	(0.433)
Chief	-0.0851		
	(0.382)		
SCC		0.252	
		(0.281)	
In/out			-0.0527
			(0.416)
Chief × Female	1.610**		
	(0.682)		
SCC × Female		0.281	
		(0.452)	
In/out × Female			0.462
			(0.662)
Constant	4.198***	4.101***	4.267***
	(0.361)	(0.238)	(0.364)
Observations	189	379	187
Adjusted R^2	0.020	0.030	0.031
Mean	4.087	4.076	4.087
SD	2.412	2.346	2.412

Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Note. The dependent variable is the number of tokens sent by the Investor. All regressions include a dummy on the Investor's gender, and each column includes a dummy for the different treatment arm (in/out, SCC, and Chief), together with the interaction of that treatment arm and the Investor's gender. The sample comes from our lab-in-the-field games conducted in Lusaka in 2017.

TABLE 2.C.8: Average return ratio by institutional treatment

	Average Return Ratio		
	(1)	(2)	(3)
Female	0.000115	-0.0427**	0.00144
	(0.0263)	(0.0200)	(0.0265)
Chief	0.0247		
	(0.0200)		
SCC		0.0340**	
		(0.0163)	
In/out			0.0218
			(0.0222)
Chief × Female	-0.0524		
	(0.0417)		
SCC × Female		0.0284	
		(0.0260)	
In/out × Female			-0.0787**
			(0.0376)
Constant	0.416***	0.418***	0.403***
	(0.0208)	(0.0143)	(0.0209)
Observations	190	380	188
Adjusted R^2	0.000	0.046	0.053
Mean	.423	.435	.423
SD	.135	.131	.135

Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Note. The dependent variable is the Trustee's average return ratio. All regressions include a dummy on the Trustee's gender, and each column includes a dummy for the different treatment arm (in/out, SCC, and Chief), together with the interaction of that treatment arm and the Trustee's gender. The sample comes from our lab-in-the-field games conducted in Lusaka in 2017.

TABLE 2.C.9: Investor's earnings

	Investor's Earnings			
	(1)	(2)	(3)	(4)
Constant	10.67***	10.67***	10.73***	10.74***
	(0.253)	(0.254)	(0.214)	(0.243)
Female	-0.0680	-0.152		
	(0.355)	(0.381)		
Chief	0.566*	0.566*	0.620**	0.607**
	(0.292)	(0.292)	(0.251)	(0.271)
SCC	1.107***	1.108***	0.914***	0.888***
	(0.282)	(0.282)	(0.221)	(0.292)
In/out	0.208	0.208	0.0577	0.0317
	(0.284)	(0.284)	(0.221)	(0.301)
Female × Chief	0.550	0.634		
	(0.657)	(0.668)		
Female × SCC	-0.588	-0.412		
	(0.429)	(0.478)		
Female × In/out	-0.256	-0.105		
	(0.418)	(0.578)		
Female × SCC × In/out		-0.306		
		(0.612)		
SCC × In/out				0.0516
				(0.442)
Observations	475	475	477	477
Adjusted R^2	0.043	0.041	0.034	0.032
Mean	10.855	10.855	10.863	10.863
SD	2.244	2.244	2.097	2.097

Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Note. The dependent variable is the Investor's final earnings. Regressors include the treatment arms (in/out, SCC, and Chief). Columns 1 and 2 additionally include a dummy variable indicating the Investor's gender, and the interaction terms of this variable with the treatment arms. The sample comes from our lab-in-the-field games conducted in Lusaka in 2017.

TABLE 2.C.10: Trustee's earnings

	Trustee's Earnings			
	(1)	(2)	(3)	(4)
Constant	6.892***	6.903***	6.805***	7.066***
	(0.538)	(0.538)	(0.488)	(0.562)
Female	-0.119	0.517		
	(0.976)	(1.176)		
Chief	-0.202	-0.204	0.114	-0.147
	(0.596)	(0.596)	(0.536)	(0.598)
SCC	-0.393	-0.394	-0.305	-0.824
	(0.512)	(0.512)	(0.432)	(0.582)
In/out	0.389	0.390	0.360	-0.164
	(0.512)	(0.512)	(0.432)	(0.677)
Female × Chief	1.400	0.771		
	(1.390)	(1.537)		
Female × SCC	0.239	-1.093		
	(0.971)	(1.338)		
Female × In/out	-0.143	-1.262		
	(0.987)	(1.464)		
Female × SCC × In/out		2.292		
		(1.657)		
SCC × In/out				1.039
				(0.864)
Observations	476	476	477	477
Adjusted R^2	-0.007	-0.005	-0.002	-0.001
Mean	6.757	6.757	6.821	6.821
SD	4.519	4.519	4.736	4.736

Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Note. The dependent variable is the Trustee's final earnings. Regressors include the treatment arms (in/out, SCC, and Chief). Columns 1 and 2 additionally include a dummy variable indicating the Investor's gender, and the interaction terms of this variable with the treatment arms. The sample comes from our lab-in-the-field games conducted in Lusaka in 2017.

3 Value Dissonance at Work

3.1 Introduction

Culture is one of the drivers of organizational performance. And yet, there is still little agreement on what culture means or what is the best empirical way to measure it. On the one hand, culture has been seen as a given set of beliefs and tacit communication codes that shape employees' and leaders' actions within an organization. Being a unique (and unspoken) code of communication among members of an organization (Cremer, 1993), this view suggests an empirical strategy based on cross firm comparison of differences in culture and associated performance (O'Reilly and Chatman, 1996; Garrett et al., 2014; Guiso et al., 2015; Erhard et al., 2009). On the other hand, studies in psychology, cognitive sciences and sociology focus on the constitutive aspects of culture (DiMaggio, 1997; Besley and Persson, 2018). These are cognitive categories, values and schema necessary for parties to think, interact and cooperate. These accounts lead to an empirical strategy that takes individuals as units of analysis and their interactions - within teams or networks - as objects of interest for performance evaluations.

Empirical papers in economics and finance have mainly adopted the former approach. This paper adopts the latter and studies the primitives of corporate culture, individual values. We move from a view of culture as a unique "latent variable" (DiMaggio, 1997) and put the complexity and multidimensionality of this concept at the core of our approach. By doing so, we are able to provide a variety of stylized facts on the way in which the sharing of different values is related to performance within an organization. By observing both employees and their direct managers, we are also able to look at the way in which organizational structures and leaders shape value sharing in the first place. This speaks to the even more difficult question of where culture comes from (Gibbons and Kaplan, 2015).

We use a proprietary dataset that maps the individual values of employees in one of the largest global banks in the world. We measure individual values using questions from the World Value Survey that ask respondents to rank the most important qualities for their children. Respondents have to choose five out of eleven values such as determination, imagination, faith, unselfishness. These questions have been validated in a number of recent papers as good proxies for the intergenerational transmission of values (see, for instance, Doepke et al., 2019).

As values are hard to observe, even in repeated interactions, perceptions of others'

values might be as important in driving behaviour as actual values. We ask respondents to list the values they think their colleagues or the bank would consider most important to transmit to children. This enables us to measure perceived misalignment between their own and colleagues' or the bank's values. We find that perceptions of misalignment are as important as actual misalignment in explaining performance at the individual level and at the team level.

Although the data come from only one firm, the sample contains geographical, hierarchical and socio-demographic variation. We first describe how values vary across these dimensions. Members of ethnic minorities tend to select different values than white or Asian groups, putting more weight on faith and hard-work than others. Gender differences are small, as well as differences between parents and non-parents. In terms of organizational roles, we find the biggest differences between people at different career bands. People at the top of the hierarchy are significantly more likely to rank imagination and determination highly, but also less likely to rank obedience, responsibility, independence and thrift among their main values compared to lower bands. These differences by hierarchical position suggest that achieving value sharing might be challenging not only horizontally, between employees, but also vertically, between employees and their managers.

We explore this by looking at the extent to which misalignment in values explains performance at both the individual and team level, either between employees or with their managers. We find that misalignment in values between an individual and his/her colleagues is negatively correlated with performance. However, on the sub-sample of employees whose manager is also a respondent, we see that vertical misalignment with the manager is more important than horizontal differences with colleagues. The crucial role of management in influencing culture has been studied in previous work. Leaders can set incentives for team members, such as rewards and punishments for norms' deviation. Other authors also stressed the key role of managers in leveraging the tacit knowledge held by employees through socialization of newcomers in the organization (Besley and Persson, 2018). "Informal management" through coordination, influence and leadership has also been put forward as a tool that relies on relational contracts among the members of an organization (Gibbons and Henderson, 2012; Gibbons and Kaplan, 2015). Our results are consistent with this latter view of what managers do. We find that values misalignment is higher in teams where communication and coordination are worse, as well as teams where the manager is overloaded with work and disagrees with senior leaders on the bank's strategy.

One way in which culture can increase performance is by allowing relationships to substitute for formal agreements through relational contracts (Chassang, 2010; Gil and Zanarone, 2016). However, the break of culture can make this harder. We see that actual as well as perceived value misalignment is correlated with lower trust in leaders, managers and colleagues. This correlation is interesting in both directions of causality. On

the one hand, limited trust might bring little experimentation on the values held by colleagues and managers, perpetuating wrong perceptions about diversity (Boisjoly et al., 2006; Beugelsdijk et al., 2019; Minni, 2019). On the other hand, value misalignment can affect agents' beliefs about others' trustworthiness, hindering incentives for cooperation (Halac, 2012; Macchiavello and Morjaria, 2015).

We next ask whether it matters what values people have and share. To reduce the dimensionality of the problem, we use latent class analysis and categorize bankers in two main types: "obedient", for those who value qualities such as obedience and faith, and "extroverts", who tend to believe in values such as imagination, self-expression and determination. Obedient types tend to have a lower performance, especially in terms of potential for career progression. Teams with a greater proportion of extrovert types do better, but this is mainly driven by the fact that these types occupy higher career bands. This suggests that sharing some specific values is more important than sharing per se' for organizational performance.

Lastly, we provide evidence on how bankers differ from common citizens by comparing their answers with World Value Survey respondents. Bankers tend to differ from average citizens on a variety of values, scoring lower on faith, thrift and obedience. "Obedient" types are closer to the average citizen, but this tends not to be rewarded by the bank. Interestingly, we find that the distance between bankers' and citizens' values decreases in those countries that were more severely hit by the financial crisis. This is suggestive evidence that aggregate industry shocks are important forces in shaping organizational culture (Besley and Persson, 2018) either through policies that try to change the values held by employees or through a change in the composition of people hired and kept by the organization.

This paper is related to two main streams of literature: corporate culture and preferences for beliefs consonance. Many papers provide theoretical models to explain the value of organizational culture. Kreps (1990) seminal model of corporate culture describes how the sharing of similar values facilitates coordination on a given organizational equilibrium. Lazear (1995) formalizes the process by which culture gets internalized by employees. Hodgson (1996) puts at the core of his work the interconnectedness between individuals and culture. Not only individuals create culture, but culture in turn molds their preferences and modes of thinking. This inspires our analysis of value dissonance at the individual level. For a review of these theoretical accounts, see Hermalin (2001). Our paper is closely related to the recent model by Besley and Persson (2018) in looking at culture and performance within the firm. They show that multiple cultures may exist for the same fundamentals, such as technology and market conditions, with unclear ranking in local performance. We also related to the work by Van den Steen (2010), who asks where culture comes from and makes explicit why the direction of causality between culture and performance is unclear.

Empirical work on corporate culture has been also growing (see Gibbons and Henderson, 2012; Gibbons and Kaplan, 2015). Anderson (1988) was an early empirical attempt to show that organizations may benefit for value congruence. He shows that organizations tend to perform better when employees have aligned values. In the same vein, Guiso et al. (2015) study integrity and find that it's positively correlated with financial performance and attractiveness of job offerings, while it is negatively correlated with the degree of unionization. In the paper, they measure integrity in terms of how workers perceive that top managers uphold this as a value. Similarly to us, their measure of culture is based on individual perceptions. We complement this work by looking at a variety of values and measuring both actual differences as well as perceptions between employees. Thakor (2016) focuses on culture in banking and posits that its value for firms can be compared with the one of financial capital. Culture in the banking sector reassure regulators of limited risk-taking and the absence of unethical behaviour, but it can also provide independent value to investors. Gartenberg et al. (2019) focus on one component of culture: employees' sharing of the same organizational purpose. In companies with strong purpose employees share a sense of the meaning and impact of their work. They find that on average purpose is not related to financial performance, but this masks heterogeneity between the types of purpose-driven organizations. Firms that are characterized by high camaraderie between workers do not perform as well as firms where managers have high clarity.

Our analysis of the relationship between value sharing and the strength of the 2008 financial crisis relates to the work by Sørensen (2002), who explores how culture responds to internal and external change. He finds that firms with strong cultures excel when encountering difficulties in volatile environments. Lins et al. (2017) examine the performance of firms engaged in Corporate Social Responsibility activities between 2008 and 2009. They show that firm specific social capital can act as insurance that pays off in moments of crisis.[1] We complement this study by looking at whether culture varies in places exposed to varying severity of the financial crisis.

This work also relates to the growing literature on preferences for belief consonance (for a review see Golman et al., 2016). Such a preference implies that people suffer a direct utility cost from knowing that others' beliefs are different from their own. There might be different microfoundations for such a taste. A common one in the literature is that conflicting beliefs threaten one's identity and reduce perceived shared identity within a group (Kahan 2010). A closely related stream of research in economics is indeed the one on identity (Akerlof and Kranton, 2005; Benabou and Tirole 2011). Taste for belief consonance has been shown to affect relationship formation (Huber and Malhotra, 2013; Mitchell et al. 2014), what media people choose (Gentzkow and Shapiro 2008, 2011) and potentially even what place they live in (Molloy et al., 2011). A few studies also looked

[1] The model by Dessein and Santos (2006) provides a framework to think about optimal firm adaptation to a changing environment.

at the effect of beliefs consonance in cooperatives, workers' associations and employees' ownership enterprises (Craig and Pencavel, 1992; Bhuyan, 2007). We contribute to this work by showing the consequences of actual and/or perceived beliefs dissonance in a company where workers do not necessarily sort on the basis of shared values.

3.2 Framework

Contract incompleteness is a pervasive force of the real world. It does not only affect the boundaries of the firm (Grossman and Hart, 1986), but also its internal organization through centralization or delegation decisions (Aghion et al., 2013). When contracts are costly or unfeasible, culture can play a crucial role in facilitating coordination, both in the short term and in repeated interactions. In this paper, we take the internal organizational structure as given and ask whether there is a positive relationship between culture and performance, both at the individual and team level.

Our first hypothesis is that, at the individual level, people whose values are further away from the average of their colleagues perform worse. In a repeated game, lower individual performance can be the consequence of the punishment suffered by a person for deviating from the team equilibrium. As the likelihood of a deviation is decreasing in the intertemporal discount factor, empirically we should expect the negative relationship between value misalignment and performance to be stronger for people who are less likely to stay in the firm. Deviations from the team equilibrium are not the only way to rationalize a negative relationship between value misalignment and individual performance. Suppose that agents face a psychological cost of not following their personal values in order to conform with the organizational culture (for instance, coming from preferences for beliefs consonance. See Golman et al., 2016). Such a cost might directly hinder individual performance, for instance by limiting cognitive functioning or bandwidth at work (Mani et al., 2013; Shah et al., 2015). In this case, the relationship between misalignment and performance would emerge even in the absence of individual deviations from a team equilibrium.

At the team level, culture can be positively correlated with performance if it allows members to coordinate on socially-maximizing equilibria. Consider a simple coordination game, such as the "battle-of-the-sexes". The game has two Nash equilibria, but standard economic theory is silent about which equilibrium we should expect to see realized (Hermalin, 2001). Culture, defined as value sharing, helps players decide or form predictions on what outcome should be achieved. Consider the following simple example. Suppose that the game is played by two team members and the two possible actions are having a pre-scheduled conference call with a client or going out for dinner (Table below). The payoffs are such that $R = S \geq s = r$.

	Call	Dinner
Call	s, R	0,0
Dinner	0,0	S, r

Both players prefer having the call/dinner together and doing one of these two actions alone will give them a zero utility. If both employees know that there is a culture of responsibility at the firm, they will easily coordinate on the first action. However, if one is responsible and the other is not, they both risk losing by choosing different actions. Perceptions can also hinder the team outcome. If the “responsible" column player mistakenly thinks that the other is not responsible, s/he would probably choose to go for dinner and at least get *r*. If this perception is wrong, both players will end up with zero, thus a lower aggregate and individual outcome. This simple case illustrates our second empirical prediction: we should expect a lower performance in teams with greater value misalignment among members.

In the given example, it’s easy to imagine that having the client’s call is a more desirable organizational outcome than going out for dinner. This implies that the sharing of those particular values which favour this outcome - such as responsibility - should be better for organizational performance. But reality is more complex. Suppose we replace the action of “having dinner" with “bringing a colleague to the emergency room". What is the dominant equilibrium for the firm in this example is less clear, with consequent uncertainty on the desirable value for employees. This is the typical situation in the “culture-as-convention" view (Hermalin, 2001); it doesn’t matter whether both players coordinate in one or the other equilibrium, as long as they’re coordinated. However, if the environment is such that one equilibrium is Pareto-dominant, a convention on the dominated equilibrium will leave money on the table. We are going to explore this issue empirically, by checking whether the positive correlation between performance and values is driven only by a few values and - if so - which ones.

3.3 Institutional context

We study values in an industry where culture has recently been under intense scrutiny: banking. The 2008 financial crisis uncovered a widespread dysfunctional culture in the sector, which regulators have since been trying to modify (Thakor, 2015). In the words of Mark Carney, the Bank of England governor (2017):

> "Multiple factors contributed to a tide of ethical drift. Market standards were poorly understood, often ignored and lacked teeth. Too many participants neither felt responsible for the system nor recognised the full impact of their actions. Bad behaviour went unchecked, proliferated and eventually became the norm."[2]

[2]Remarks at the Banking Standards Board Panel “Worthy of trust? Law, ethics and culture in banking” by Mark Carney, Governor of the Bank of England, 21 March 2017

A common narrative of the financial crisis is that coordination on a bad equilibrium, which reduced welfare for society, could be sustained through inflated individual monetary incentives and value systems detached from society. Andrew Haldane, from the Bank of England, effectively expresses this view (2012):

> "The crisis was the story of a system with in-built incentives for self-harm. Avoiding those self-destructive tendencies means changing the incentives and culture of finance, root and branch."[3]

A survey by Deloitte in 2013 found that 65% of interviewed senior leaders believed there were significant cultural problems in the industry. Interestingly, they also thought that the problems were less extensive in their own banks than in others.

Since 2008, huge investments have been made to try and change the culture in banking. For instance, in 2015 the U.K. banking sector created the Banking Standards Board (BSB), a body meant to promote good practice among banks and building societies. In the Netherlands, bank leaders must now swear an oath to put the customers' interests first.[4] In the U.S., Goldman Sachs reacted to a 550 USD millions fine by the Securities and Exchange Commission (SEC) by launching a fundamental review of its practices and culture, published in 2011. The review consists of 39 recommendations and includes compulsory courses on ethics for senior managers (The Economist Intelligence Unit, 2013). Some authors posit that a strong banking culture can be considered a complement to financial capital on the bank's balance sheets (Thakor, 2016).

We believe that a snapshot of bankers' values ten years after the crisis, and their distance with respect to common citizens, can add insights to these and other investments in changing the banking culture.

3.4 Data

The data we use in this paper come from a partnership with one of the world's largest banks. In May 2017, we collected data on employees' values and their perceptions of values held by the bank and their colleagues. We chose our main questions from the World Value Survey (WVS from hereon). Here are the questions we asked:

1. Here is a list of qualities that children can be encouraged to learn at home. Which do you consider to be especially important? (rank the top 5)

2. Imagine [BANK NAME] had a family. What qualities do you think that [BANK NAME] would consider especially important for its children to learn at home? (select the top 3)

[3] Speech to Occupy Economics "Socially useful banking" by A. G. Haldane, Executive Director, Financial Stability, Bank of England, 29 October 2012.

[4] See the article by Zaring (2017) on the New York Times at this link.

3. Now thinking about your colleagues where you work, what qualities do you think that they would consider especially important for children to learn at home? (select the top 3)

For each of these questions, respondents had to choose among the following eleven values: Independence, Hard work, Feeling of responsibility, Imagination, Tolerance and respect for other people, Thrift and saving money/things, Determination and perseverance, Religious faith, Unselfishness, Obedience, Self-expression.

Our measures have been shown to correlate with the intergenerational transmission of values in a few recent papers. Guiso, Sapienza and Zingales (2003) use them to study the link between religion and preference for thriftiness, as measured in the WVS. In a study on the relationship between trust and regulation, Aghion et al (2010) use the WVS measures of desired children's values as proxies for family civic education. In a similar vein, Bauer et al (2014) use these variables to study the role of children's socialization in other-regarding preferences. More recently, Doepke and Zilibotti (2017) and Doepke et al. (2019) use these measures as proxies for different parenting styles (e.g., authoritarian, relaxed) and show that they are correlated with a variety of country-level macroeconomic conditions such as income inequality, the return to education, and redistributive policies. We believe that these questions can be a good proxy for individual values in our context and should be relatively free from social desirability bias (Fischer, 1993). Despite being part of a survey sent by the employer, there is no clear answer which respondents could select to project a favourable image of themselves.

The questions were added to the organization's annual Census sent to all the employees over the world. The typical survey response rate is 40 percent, which leads to 95,138 respondents for the full survey. The survey collects information on employees' engagement with the bank's strategic priorities as well as general sentiments about the workplace. Our questions on individual values were randomly displayed to 40 percent of the basis to reduce average survey duration. The total number of respondents to our values questions is 38,827 across 55 countries. The percentage of randomly selected respondents across countries varies from 26% to 67%. Table 3.1 shows summary statistics of the sample of survey respondents, divided by whether they were selected to reply to the values questions or not.

Our core question is whether value sharing between employees enables better performance. In the absence of an exogenous variation that can be used to identify causal impacts, we present some stylized facts on the correlation between our measures of values and proxies for performance. To build these proxies, we merged the survey data with personnel records and use three annual measures of individual performance. The first is the supervisors' assessment of employees' performance relative to expectations for that role. For most of our sample, this measure is based on financial performance indicators of the profitability and growth of employees' client portfolios. The second measure is

the supervisors' assessment of whether the employee has "good behaviours" such as integrity, cooperation and connection with the customer. Relationships with customers are particularly important for the 39% of our sample who are in client facing roles. Both the behaviour and financial performance measures are those used to determine the annual bonuses. Our third measure is the supervisors' assessment of the potential for vertical progression within the bank in the near future. The three measures are on a scale between 1 (min) and 4 (max). Throughout the paper we standardize these measures to be mean zero and with unit standard deviation.

3.5 Values across the organization

The top panel of Figure 3.1 provides an overview of bankers' values. In aggregate, the top five values are responsibility, tolerance, independence, determination and hard work, all mentioned by more than 50% of respondents. The bottom three values are faith, obedience and thrift, mentioned by less than 20%. Following Doepke, Sorrenti and Zilibotti's (2019) classification of parenting styles, bankers' seem to rank *relaxed* values such as independence and imagination more highly (in 4^{th} and 6^{th} position respectively) as compared to *intensive* values such as hard-work and obedience (in the 5^{th} and 10^{th} position respectively). This is true even when we compare bankers with the general population from the WVS (Figure 3.1 bottom panel). Bank employees are significantly less likely to mention faith, thrift and obedience than the general population, but more likely to mention imagination, self-expression and determination. The figure looks extremely similar if we limit WVS respondents to those with an educational level comparable to bankers (i.e. those who have completed high school or above). Within country comparisons reveal a similar difference between bankers and WVS respondents: obedience, thrift, faith and unselfishness are among the values mentioned less often by bankers than common citizens. Figure 3.A.1 shows the comparison across continents.

3.5.1 Values by demographics and organizational role

Table 3.1 shows that the bank workforce is quite heterogeneous in terms of gender, ethnicity and age. These are all characteristics which might correlate with different stated values. In Figure 3.2, we show differences in the most important values listed by these different demographic groups.

In the top figure, each bar gives the percentage point difference in the proportion of people who mentioned a given value between members and non-members of an ethnic minority.[5] Differences tend to be small, within 10 p.p. in absolute value, but they are spread over a wide range of values. The biggest difference is observed in faith; a member

[5]Members of a minority are defined as people who answer "yes" to the question: "Do you consider that you belong to an ethnic minority?". Most people of Black, Latino or Mixed ethnicity answer "yes" to this question, while only 2% of White people and 30% of Asian people say "yes".

of a minority is 10 p.p. more likely to mention this value among their top five than someone who isn't a member of a minority. This difference represents a twofold increase over the aggregate proportion of 10% of workers who choose faith among their top ranked values. Even smaller differences are observed in the bottom panel by parenting status. This fact gives reassurance on the validity of these measures, where answers do not greatly differ depending on parenthood. The graph in the middle shows differences by gender. They tend to be concentrated in a few values: imagination and hard-work, which women are less likely to mention, and responsibility, tolerance and independence, which women are significantly more likely to mention.

The main takeaway of this figure is that diversity based on observables also brings diversity in values. This could be a resource as well as a challenge for building organizational culture. Nevertheless, the way in which such diversity translates into culture ultimately depends on team composition, horizontally, and on the roles covered by different groups, vertically. This suggests looking at whether there are differences in values between people who occupy different organizational roles.

Figure 3.3 shows that there are small differences in value rankings by tenure or customer facing vs back office employees. There is almost no difference in stated values between people with five to nine years of tenure and those who have been at the bank for four years or less. The former group tend to be less likely to mention hard work and more likely to mention tolerance, faith and responsibility among their top values, but differences are all smaller than 2.5 p.p.. Senior employees with more than ten years of work in the bank are more likely to mention determination, unselfishness and tolerance, but less likely to mention obedience, self-expression and imagination than junior employees of any tenure.

In contrast, the central panel of Figure 3.3 shows that there are substantial differences in stated values depending on the hierarchical position held. We classified employees in three career bands: top (0.8%), middle (35%) and bottom (64%). The central panel of Figure 3.3 shows differences in the proportion who mentioned each value between the top (in blue) and the middle (in red) with the bottom band. People at the top of the hierarchy are significantly more likely to rank imagination and determination highly as compared to both the middle and bottom bands. Both of these are values correlated with high aspirations and ambitions. The top band is less likely to rank obedience, responsibility, independence and thrift among their main values than both the middle and bottom bands. These differences by hierarchical position suggest that achieving the sharing of common values might be challenging not only horizontally, between employees, but also vertically between employees and their managers. The implications for performance might be very different in these two cases, especially as the manager has actual power to award promotions and bonuses. Results in the next Section will confirm that vertical differences seem to matter to a greater extent than horizontal ones.

3.6 Value misalignment

3.6.1 Measures

We measure the value of culture by exploiting variation in the (dis)similarity in values listed by employees in our survey questions. We share the drawbacks of a self-reported measure with many studies in corporate culture (Guiso et al., 2015; Graham et al., 2019). Differently from most of the studies in this area, we ask about personal values which are not directly related to the workplace. This is both an advantage and disadvantage of our strategy. On the one hand, it limits demand bias. On the other, these values might be less relevant drivers of behaviour in the workplace.

Questions 2 and 3 of Section 3.4 provide direct measures of *perceived* value misalignment with colleagues and the bank. We construct measures of *actual* value misalignment with colleagues by comparing individual answers to question 1 with the answer given by their colleagues. This proxy is available for the smaller sample of teams that contains multiple responses to the survey. The distinction between perceived and actual value misalignment has important implications for organizations interested in cultural change. Perceptions can be changed through better coordination and communication, while actual differences can potentially only be tackled through different selection into the firm.[6] Multiple observations within a team also enable us to construct team-level measures such as average actual and perceived misalignment between team members and/or the manager, and between team members and the bank.

A complication in constructing indexes of value misalignment in our data is that we only asked respondents for partial rankings. We overcome this by using mathematical methods that compute the average distance between top-k lists, defined as rankings where only the top k members of an ordering are observed. We use two main methods: Absolute Spearman ρ, which is an adaptation of Spearman's footrule, and Kendall τ. The following paragraphs provide a brief description of each of them. We are going to use the Absolute Spearman ρ index throughout the paper. For robustness, we repeat the main tables using the Kendall τ index and include them in the Appendix.[7]

Absolute Spearman ρ

For each pair of people in a team, we subtract person 1's ranking of a given value from person 2's ranking. We do this for all eleven values and sum up the absolute value of each difference to get the distance. Every value which is not ranked among the top k is given rank $k + 1$:

[6] For a provoking view on culture-based recruitment, see McCord (2018).

[7] We use both methods to construct either actual or perceived misalignment. To construct perceived misalignment we need to further adapt these measures. Respondents were indeed asked to report only the top three values for their colleagues and the bank, without specifying how to rank them. We compare the individual's top three values with the ones reported for their colleagues or the bank applying the rules specified in this Section.

$$\rho_{abs}(R_1, R_2) = \sum_{\{i=1\} \in N} |r_1(i) - r_2(i)|$$

where i is a value, $r_j(i)$ is the ranking for person $j \in \{1, 2\}$ and N is the set of eleven values.

Kendall τ

For each pair of people in a team, we calculate the distance between their rankings of each possible pair of values and sum them up over all pairs. Let $N = \{1, ...n\}$ be the set of value pairs and R_1, R_2 be the two top k rankings on N. The Kendall τ distance is defined as:

$$d_{Kendall}(R_1, R_2) = \sum_{\{i,j\} \in N} K_{i,j}(R_1, R_2)$$

where:

- $K_{i,j}(R_1, R_2) = 0$ if i and j appear in the same order in R_1 and R_2
- $K_{i,j}(R_1, R_2) = 1$ if i and j appear in the opposite order in R_1 and R_2
- $K_{i,j}(R_1, R_2) = \frac{1}{2}$ if both i and j appear in position $k + 1$ in a ranking (i.e. not in the top k positions) and in positions ahead of $k + 1$ in the other ranking

For example, take a given pair of values: independence (i) and obedience (j). If both people rank independence and obedience among their top five values, distance will be zero or one depending on whether the values are ranked in the same or different order. Now consider the case in which one person ranks both independence and obedience, but the other ranks only one. If the unranked value in the second list is the same as the lower ranked value in the first, distance is zero. If the unranked value in the second list is the same as the higher ranked value in the first, distance is one. If one person ranks independence, but not obedience, and the other ranks obedience, but not independence, the distance is always one. Finally, if one person ranks independence and/or obedience, but the other ranks neither, distance is a half. The penalty is less than one because we don't know how the person who doesn't rank either value would order them.

3.6.2 Value misalignment and performance

We measure the relationship between value misalignment and performance using the following specification:

$$y_{ic} = \alpha + \beta Misalignment_{ic} + X'_{ic}\delta + \lambda_c + \epsilon_{ic}$$

where $Misalignment_{ic}$ is one of the Absolute Spearman ρ indexes of actual or perceived misalignment described in the previous section for person i (or team) in country c. y_{ic} is a standardized measure of financial, behavioural or potential performance, λ_c are country fixed effects and X_i includes controls for job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.[8]

Table 3.2 shows the main stylized fact of the paper. Greater value misalignment is correlated with lower individual and team performance. The dependent variable is financial performance in Columns (1) and (2), behavioural performance in (3) and (4) and potential for progression in (5) and (6). At the individual level, the effect size is small but precisely estimated: a standard deviation increase in the absolute Spearman ρ index is correlated with 2.6% of a standard deviation decrease in either financial or behavioural performance. Panel B of Table 3.2 shows a similar story at the team level, but the relationship between actual value differences and performance has a smaller magnitude of around one quarter of the effect for individuals. Appendix Table 3.A.6 checks for robustness using the Kendall τ index of misalignment.

The score for potential rating is an exception. Columns (5) and (6) show that perceived misalignment with the bank has a positive relationship with the manager's assessment of the potential for vertical career progression. This is consistent with the evidence of large gaps in top values between hierarchical levels. If values drive behaviour, misalignment with the bank might be correlated with rare actions which are deemed to be required among the leadership.[9] One caveat is that the "potential" score is available for only half of the sample.[10]

Somewhat surprisingly, the magnitude of the correlations of Table 3.2 is similar for perceived and actual misalignment, despite their raw correlation being only around 0.10. This suggests that different processes might mediate the relationship between performance and perceptions, on the one hand, and actual differences in values, on the other.

Figure 3.5 checks whether the negative correlation in performance is driven by a few particular values. Misalignment seems to be equally bad for performance across all the listed values. None of the coefficients of a regression of financial or behavioural performance on the eleven value misalignment indexes is positive and significant. The only coefficient which is positive and non-negligible in magnitude is the one on determination, which we saw was mentioned disproportionately by higher career bands (Figure 3).

Table 3.3 explores whether the negative relationship between actual misalignment

[8] Assignment is equal to zero if the role covered by the person is the first in the bank, one if it's the second role held in the bank etc.

[9] Estimations are robust to the inclusion of a full set of controls for career bands.

[10] Limiting the sample to only the portion with potential rating does not affect the sign and magnitude of the negative relationship between financial or behavioural performance and misalignment.

and performance is driven by horizontal differences between colleagues or vertical differences with managers. This exercise can be performed only on the sample of respondents whose manager also replied to the values questions, implying a reduced sample size. In Columns (1) to (3), the dependent variable is financial performance, in (4) to (6) it is behavioural performance and in (7) to (9) it is potential for progression. Columns (1), (4) and (7) repeat the specification of Table 3.2 in this smaller sample for benchmarking. Columns (2) and (4) suggest that the negative correlation between actual misalignment and performance is driven by differences in values with managers and not with colleagues. Nevertheless, Columns (3) and (6) show that the negative correlation between perceived misalignment with colleagues or the bank and performance is not only driven by actual differences with the manager.

One mechanism through which value misalignment can be negatively correlated with performance is by limiting trust and thus the scope for relational contracting (Chassang, 2010; Gil and Zanarone, 2016). Table 3.4 looks at this channel and correlates misalignment indexes with self-reported trust in colleagues, managers, and leaders (Columns (1) to (3)). Perceived misalignment in both the bank and colleagues is highly correlated with all the self-reported measures of trust. Actual misalignment with colleagues is negatively correlated with both trust in colleagues, leaders and feelings of being trusted. This correlation is interesting in both directions of causality. On the one hand, limited trust might bring little experimentation on the values held by colleagues and managers, perpetuating wrong perceptions about diversity (Boisjoly et al., 2006; Beugelsdijk et al., 2019). On the other hand, value misalignment can affect agents' beliefs about others' trustworthiness, hindering incentives for cooperation (Halac, 2012; Macchiavello and Morjaria, 2015). Appendix Table 3.A.7 repeats the regressions on trust using the Kendall τ index of value misalignment.

3.6.3 Where does misalignment come from?

The previous Section hints at the crucial role of manager-employee relationships for organizational performance. In this Section we dig deeper into the role of leaders in fostering or depleting culture. A rich literature in economics (Bandiera et al., 2019; Gibbons and Henderson, 2012; Hermalin, 2012; Bertrand and Schoar, 2003) and management (Mintzberg, 1973; Starr-Glass, 2019) studies the role of leaders for organizational performance. In relation to corporate culture, leaders can help agents' coordination by facilitating communication flows (Dewan and Myatt, 2008) and setting good examples in an informal way (Gibbons and Kaplan, 2015). Alternatively, they might set a mission for the supervised team (Wendt et al., 2009; Bolton et al., 2012).

If we take team composition as fixed, leaders might affect value alignment by influencing the extent to which employees get to know each others' values, for instance through different designs of processes and delegation. In this case, we should expect perceptions of alignment to be correlated with different managers' characteristics more

than actual misalignment, which should be orthogonal to managers' choices. This is the relevant case in our setting; direct managers have limited power in affecting team composition, which is decided at higher levels of leadership.

If we allow team composition to vary endogenously with the manager's appointment, value alignment can be affected by employees' selection as well. In this case, perceptions and actual misalignment could both be related to the manager. In both cases, we expect managers' attitudes to matter in the formation of team-level culture.

Table 3.5 correlates measures of misalignment in values at the team level with managers' attitudes towards colleagues, the organization and their own role. In Columns (1) to (3) the dependent variables are managers' opinion of their productivity: whether they think there is not enough time to do things properly, whether they feel that conditions in the job do not allow them to be as productive as possible and whether they are proud of their work. In Columns (4) to (6) the dependent variables are managers' confidence in the future, disagreement with leaders and confidence in colleagues. In Column (7) the dependent variable is managers' perceived misalignment in values with the bank.

Table 3.5 shows three main results. First, managers' attitudes towards colleagues, the organization and their own role are correlated with perceptions of value misalignment in their teams. However, managers' attitudes are not related to actual alignment between employees, suggesting that the assumption of exogenous team composition is appropriate. Secondly, managers' attitudes are correlated with perceptions of the bank's overall values, but not of colleagues' values. This suggests that managers might play a key role in spreading aggregate culture locally to their teams, as highlighted in Besley and Persson (2018). Third, the correlation between managers' attitudes and value misalignment is small in magnitude. For instance, managers' disagreements with bank leaders increases team-level misalignment with the bank by 1% of a standard deviation. Nevertheless, such a small difference might have bigger implications in terms of actual team behaviour.

What do managers do in teams with better culture? Managers might socialize team members into their own views directly or indirectly through the team organization design. In a long term equilibrium we should expect organizational culture to help managers reduce coordination and communication costs needed to motivate team members. However, in a dynamic view, managers might need to invest time and effort to create a stable team culture. Table 3.6 shows that better communication and coordination strategies are indeed positively correlated with team culture. Our communication index is constructed by averaging employees' agreements to statements such as "Where I work, issues of right and wrong are discussed in team meetings" or " Where I work, everyone has the chance to say what they think about issues". Our coordination index averages variables such as the frequency of team meetings and check-in with managers, whether the manager gives feedback on employees' behaviour or the bank's strategy. Both the creation of forums of discussion as well as the absence of negative consequences for expressing personal opinions seems to matter for reducing perceived misalignment among

employees.

3.7 Does it matter what values people have and share?

Are certain values rewarded disproportionately by the bank? Are these values driving the relationship between value misalignment and performance? Figures 3.4 and 3.5 report the coefficients of the following individual-level specification:

$$y_{ic} = \alpha + \sum_{j=1}^{11} \beta_j Value_{icj} + X'_{ic}\delta + \lambda_c + \epsilon_{ic}$$

where $Value_{icj}$ is an indicator variable for whether person i in country c mentioned a certain value j among her top five in Figure 3.4. $Value_{icj}$ is the Absolute Spearman ρ index of misalignment in value j in Figure 3.5. y_{ic} is the standardized measure of financial or behavioural performance, λ_c are country fixed effects and X_{ic} includes controls as previously defined. In Figure 3.4, the omitted category is mentioning "responsibility", which is the most common value mentioned across the bank.

Figure 3.4 shows that there is no single value which is disproportionately rewarded by managers' assessments. Determination and tolerance are the only values with a positive point estimate, but they are both not significant at the 95% confidence level. Compared to responsibility, most values tend to have a negative correlation with performance. The biggest negative differences are in faith, unselfishness, thrift, obedience and imagination, especially in relation to financial performance. Importantly, these are also the values on which bankers differ the most with society at large (see Figure 3.1). Evidence in Figure 3.5 confirms this view using the misalignment index in each value.

Both omitted variables and reverse causality might impact these estimates. The latter problem seems to be unlikely in this context, as it's unclear how performance might affect people's likelihood of reporting one particular value out of the ones listed. However, unobservable individual characteristics might be correlated with both values and performance, driving the relationship shown in Figure 3.4.

To make progress on this aspect and reduce the dimensionality of the problem, we use latent class analysis (LCA) to categorize respondents into types. Latent Class Analysis is a type of structural equation modelling used to find groups of cases in multivariate categorical data. These groups are called "latent classes". In our context, LCA helps to find classes of employees who share similar values and whose values are independent, conditional on class belonging. Such a categorization allows us to measure value sharing as the proportion of employees of a certain type within a team.

Table 3.7 shows the proportion of people who mentioned each of the eleven values in each of the two identified latent classes. Values are ranked from the top to the bottom of the table in order of absolute difference between employee type one and two. The top values mentioned by type one are imagination, self expression, independence and

determination. We will label this type as “extrovert". In contrast, type two employees are more likely to mention values such as obedience, faith, thrift and hard-work. We will label them “obedient" types. Notice that this set of values implies that the "obedient" type is closer to common citizens as compared to the average banker (as shown in Figure 3.1). Appendix Table 3.A.3 shows some differences in demographic characteristics between the two types. Overall such differences are small in magnitude, but interesting. For instance, members of ethnic minorities are 8% of the extrovert group and 11% of the obedient group. People in customer facing roles are more frequent among the obedient types (40% vs 37%) and top career bands are instead more frequent among the extroverts. The last rows in Table 3.A.3 show that there are significant differences in performance ratings between the two groups. We confirm this in a regression framework, where we can control for the various differences in socio-demographic variables between the two groups of people. Table 3.8 shows that extrovert employees have a significantly higher performance than obedient types in all the measures considered. This translates into higher team performance when extrovert people are the majority. The first row of Table 3.9 shows that teams with a more diverse composition tend to perform worse than homogeneous teams when either extroverts or obedients are the majority. However, this relationship is entirely driven by teams with more than 60% of extrovert employees, as shown in the last two rows of the Table. This evidence does not simply mean that having more high-achievers in a team increases team performance. It also implies that a diverse team is not better than a team full of low achievers. Value sharing thus seems to bring some advantages in terms of performance even when the shared values are not the best ones.

In the regressions of Table 3.9 we control for the team proportion of people from the top career band. If we add controls for the proportion of people in the middle or bottom career bands, the coefficients on the proportion of extroverts become smaller and non-significant, with the exception of Columns (5) and (6) on potential rating. This interestingly suggests that the relationship between extrovert values and performance is mediated by different proportions of these values across hierarchies.

There are two main takeaways from this exercise. First, while value sharing is positively correlated with performance, banking seems to be rewarding some values more than others, especially in terms of career progression. Secondly, the most rewarded values are also the ones with the largest distance from common citizens. Unfortunately, we cannot say whether people who are more similar to WVS respondents are simply worse bankers, whether their values are not appreciated by the bank or whether the coordination achieved by these values brings only modest outcomes. However, this is an intriguing piece of evidence for thinking about the sector a decade after the last financial crisis. Are bankers able to have clients' and societal interests at the top of their minds despite their cultural distance with the average citizen? In the last section, we conclude by looking at cross-country differences in misalignment within the bank and distance with

WVS respondents.

3.8 Industry shocks and organizational values

We merged our data with data on the change in unemployment rates between 2007 and 2009, used by previous work as a measure of severity of the financial crisis (Stevenson and Wolfers, 2011). For each country, we then compute the average distance in values between bankers and common citizens. We first compute the difference between the proportion of bankers and the proportion of WVS respondents that mentioned a certain value among their top five. We then average these differences across the eleven values. Figure 3.6 shows that there is a negative relationship between change in unemployment between 2007 and 2009 and distance between bankers and WVS respondents. This seems against the common narrative of the crisis generated by bankers who are detached from society. However, employment in banking is not randomly assigned. What Figure 3.6 tells us is that those bankers who decided to stay in the sector after 2007 - or were recently hired - have values which are more similar to common citizens in those countries which were more heavily affected by the crisis. There might be different explanations for this pattern. The experience of the crisis might have changed the types of individuals who sort into the sector, decided to stay or who are hired by the bank.

Figure 3.7 shows the average distance between bankers' and WVS values by terciles of change in unemployment rates between 2007 and 2009 and by tenure at the bank. We can see that, in both the second and third tercile of unemployment changes, the increased similarity between bankers and common citizens is driven by cohorts who started working in the bank after 2000. A mix of selected stayers and new hires might be driving the relationship of Figure 3.6. In the first tercile, both people with the highest and lowest tenure have values which are further away from the WVS.

This is suggestive evidence that aggregate industry shocks are important forces in shaping organizational culture (Besley and Persson, 2018). Shocks can trigger abrupt policies that try to change the values held by employees (see Section 3.3) or require a change in the composition of people hired and kept by an organization. Despite the fast pace in which these policies can happen, changing organizational culture inevitably takes time. The financial crisis might have encouraged banks to either review internal values or hire new types of people more similar to common citizens. However, a decade might not be enough to fully converge on a new culture. Figure 3.8 indeed shows that misalignment between employees within the bank is increasing in the severity of the financial crisis. This is a subtle way in which aggregate shocks can have medium to long term consequences on organizational performance.

3.9 Concluding remarks

As high-skilled workers increase their time at work, the workplace becomes the main site where individual values shape behaviour and relationships. Dissonance between personal values and the ones held by colleagues and/or the organization can negatively affect individual as well as organizational outcomes. Understanding when such dissonance emerges and its implications for organizational performance has important implications for employees' motivation, selection and organizational design.

In this paper we provided evidence that the sharing of individual values matters for employee and team performance within a multinational firm. Both perceptions and actual value alignment matter. As highlighted in previous work, we find that managers have a crucial role in shaping employees' perceptions of shared values through both personal attitudes and team practices. Organizational initiatives which foster team cohesion and increase familiarity between managers and employees might be beneficial in reducing perceived value misalignment, but it's unclear whether their benefits would be high enough to exceed their costs.

Management is an internal force which shapes individual culture. We also find evidence that external forces, such as aggregate industry shocks, can increase the challenge of achieving a uniform organizational culture and have long-term consequences through their impact on employees' value sharing. Alternatively, employees' value sharing could be a shock-mitigating force that provides insurance on team productivity in bad times (Aghion et al., 2017). Such an interplay between industry-wide economic conditions and organizational values is an exciting venue for follow-up research. Another important next step in this work is to move towards an identification of the causal effect of value misalignment on employees' performance.

This paper studied the micro-foundations of corporate culture at a certain point in time and within a given organizational structure. This offers only a limited perspective on the dynamics of organizational change. Organizations constantly evolve as a result of both internal and external pressures for change. The aggregation of values in larger collectivities and their evolution over time as the composition of workers change are exciting topics left for future research.

3.10 References

Aghion, P., Algan, Y., Cahuc, P., & Shleifer, A. (2010). Regulation and distrust.*The Quarterly Journal of Economics*, 125(3), 1015-1049.

Aghion, P., Bloom, N., & Van Reenen, J. (2013). Incomplete contracts and the internal organization of firms. *The Journal of Law, Economics, & Organization*, 30(1), i37-i63.

Aghion, P., Bloom, N., Lucking, B., Sadun, R., & Van Reenen, J. (2017). *Turbulence, firm decentralization and growth in bad times*. NBER working paper w23354. National Bureau of Economic Research.

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

Anderson, E. (1988).Transaction Costs as Determinants of Opportunism in Integrated and Independent Sales Forces. *Journal of Economic Behavior and Organization*, 9, 247–264.

Bénabou, R., & Tirole, J. (2011). Identity, Morals and Taboos: Beliefs as Assets. *Quarterly Journal of Economics*, 126(2), 805–855.

Baker, G., Gibbons, R., & Murphy, K. J. (2002). Relational Contracts and the Theory of the Firm. *The Quarterly Journal of Economics*, 117(1), 39-84.

Bandiera, O., Prat, A. Hansen, S. & Sadun, R. (2019). CEO Behavior and Firm Performance. *Journal of Political Economy*, Forthcoming.

Bauer, M., Chytilová, J., & Pertold-Gebicka, B. (2014). Parental background and other-regarding preferences in children. *Experimental Economics*, 17(1), 24-46.

Bertrand, M., & Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly journal of economics*, 118(4), 1169-1208.

Besley, T., & Persson, T. (2018). The Joint Dynamics of Organizational Culture, Design, and Performance. *Unpublished manuscript*.

Beugelsdijk, S., Klasing, M. J. & Milionis, P. (2019). Value Diversity and Regional Economic Development. *Scandinavian Journal of Economics*, 121, 153-181.

Bhuyan, S. (2007). People' Factor in Cooperatives: An Analysis of Members' Attitudes and Behavior. *Canadian Journal of Agricultural Economics*, 55(3), 275–298.

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M. & Eccles, J. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review*, 96(5), 1890-1905.

Bolton, P., Brunnermeier, M. K., & Veldkamp, L. (2012). Leadership, coordination, and corporate culture. *Review of Economic Studies*, 80(2), 512-537.

Chassang, S. (2010). Building routines: Learning, cooperation, and the dynamics of incomplete relational contracts. *American Economic Review*, 100(1), 448-65.

Craig, B., & Pencavel, J. (1992). The Behavior of Worker Cooperatives: The Plywood Companies of the Pacific Northwest. *American Economic Review*, 82(5), 1083-1105.

Cremer, J. (1993). Corporate culture and shared knowledge. *Industrial and Corporate Change*, 2, 351–386.

Deloitte (2013). *Culture in banking. Under the microscope.* The Deloitte Bank Survey

Dessein, W., & Santos, T. (2006). Adaptive organizations. *Journal of Political Economy*, 114(5), 956-995.

Dewan, T., & Myatt, D. (2008). The Qualities of Leadership: Direction, Communication, and Obfuscation. *American Political Science Review*, 102, 351–368.

DiMaggio, P. (1997). Culture and cognition. *Annual review of sociology*, 23(1), 263-287.

Doepke, M., & Zilibotti, F. (2017). Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, 85(5), 1331-1371.

Doepke, M., Sorrenti, G., & Zilibotti, F. (2019). The economics of parenting. *Annual Review of Economics*, 11, 55-84.

Erhard, W.H., Jensen, M.C. & Zaffron, S. (2009). *Integrity: a positive model that incorporates the normative phenomena of morality, ethics, and legality*. Harvard Business School NOM Working Paper No. 06-11.

Fisher, R. J. (1993). Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2), 303–315.

Garret, J., Hoitash, R. & Prawitt, D. F. (2014). Trust and Financial Reporting Quality. *Journal of Accounting Research*, 52, 1087-1125.

Gartenberg, C.M., Prat, A. &. Serafeim, G. (2019). Corporate Purpose and Financial Performance. *Organization Science*, 30 (1), 1-18.

Gentzkow, M., & Shapiro, J. M. (2008). Competition and Truth in the Market for News. *Journal of Economic Perspectives*, 22(2), 133–154.

Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *Quarterly Journal of Economics*, 126(4), 1799–1839.

Gibbons, R. & Henderson, R. (2012). What do managers do? Exploring persistent performance differences among seemingly similar enterprises. In R. Gibbons and J. Roberts (eds.). (2012). *Handbook of organizational economics*, (pp. 680–731). Princeton University Press.

Gibbons, R., & Kaplan, R. S. (2015). Formal Measures in Informal Management: Can a Balanced Scorecard Change a Culture? *American Economic Review*, 105 (5), 447-51.

Gil, R. & Zanarone, G. (2016). New Frontiers in Empirical Research on Informal Contracting. *Journal of Institutional and Theoretical Economics*, 172(2), 390-407.

Golman, R., Loewenstein, G. F., Moene, K. O., & Zarri, L. (2016). The Preference for Belief Consonance. *Journal of Economic Perspectives*, 30(3), 165–188.

Gordon, G. & DiTomaso, N. (1992). Predicting Corporate Performance from Organizational Culture. *Journal of Management Studies*, 29(6), 783-798.

Graham, J. R., Harvey, C. R., Popadak, J., & Rajgopal, S. (2017). *Corporate culture: Evidence from the field.* NBER working paper w23255. National Bureau of Economic Research.

Grossman, S. J., & Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4), 691-719.

Guiso, L., Sapienza, P., & Zingales, L. (2003). People's opium? Religion and economic attitudes.*Journal of Monetary Economics*, 50(1), 225-282.

Guiso, L., Sapienza, P., & Zingales, L. (2015). The value of corporate culture. *Journal of Financial Economics*, 117, 60-76.

Halac, M. (2012). Relational Contracts and the Value of Relationships. *American Economic Review*, 102(2), 750-779.

Hermalin, B. (2001). Economics and Corporate Culture. In C. L. Cooper, S. Cartwright, and P. C. Earley (eds.). *The International Handbook of Organizational Culture and Climate*, (chapter 10). New York: John Wiley & Sons.

Hermalin, B. E. (2012). Leadership and corporate culture. In R. Gibbons and J. Roberts (eds.). (2012). *Handbook of organizational economics*, (pp. 432-478). Princeton University Press.

Hodgson, G. M. (1996). Corporate Culture and the Nature of the Firm. In J. Groenewegen, (ed.). *Transaction Cost Economics and Beyond*, (pp. 249-269). Springer, Dordrecht.

Huber, G., & Malhotra, N. (2013). *Dimensions of Political Homophily: Isolating Choice Homophily along Political Characteristics*. Working Paper 3108, Stanford University.

Kahan, D. (2010). Fixing the Communications Failure. *Nature*, 463(7279), 296–297.

Kreps, D. M. (1990). Corporate Culture and Economic Theory. In J. E. Alt and K. A. Shepsle, (eds.). *Perspectives on Positive Political Economy*, (pp. 90-143). Cambridge, England: Cambridge University Press.

Lazear, E. P. (1995). Corporate Culture and the Diffusion of Values. In H. Siebert *Trends in Business Organization*, (pp. 89-133). Tubingen, Germany: J.C.B. Mohr (Paul Siebeck).

Lins, K.V., Servaes, H. & Tamayo, A. (2017), Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. *The Journal of Finance*, 72, 1785-1824.

Macchiavello, R., & A. Morjaria (2015). The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports. *American Economic Review*, 105(9), 2911–2945.

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, 341(6149), 976-980.

McCord, P. (2018). How to Hire. *Harvard Business Review*, 96(1), 90-97.

Minni, V. (2019). Cultural Distance And Performance: Evidence From A Multinational Firm. *Unpublished manuscript*.

Mintzberg, H. (1973). Strategy-making in three modes. *California management review*, 16(2), 44-53.

Mitchell, A., Gottfried, G., Kiley, J., & Matsa, K. E. (2014). *Political Polarization and Media Habits*. Pew Research Center.

Molloy, R., Smith, C. L. & Wozniak, A. (2011). Internal Migration in the United States. *Journal of Economic Perspectives*, 25(3), 173–96.

O'Reilly, C. A., & Chatman, J. A. (1996). Culture as social control: Corporations, cults, and commitment. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior: An annual series of analytical essays and critical reviews*, (vol. 18, pp. 157-200). US: Elsevier Science/JAI Press.

Sørensen, J. (2002). The Strength of Corporate Culture and the Reliability of Firm Performance. *Administrative Science Quarterly*, 47(1), 70-91.

Shah, A. K., Shafir, E., & Mullainathan, S. (2015). Scarcity frames value. *Psychological Science*, 26(4), 402-412.

Starr-Glass, D. (2019). The Culture of Management and the Management of Culture: An Introduction. In: Machado C., Davim J. (eds) *Management Science. Management and Industrial Engineering*, (pp. 49-74). Springer, Cham.

Stevenson, B. & Wolfers, J. (2011). Trust in Public Institutions over the Business Cycle. *American Economic Review: Papers and Proceedings*, 101(3), 281–287.

Thakor, A. (2015). The financial crisis of 2007–2009: Why did it happen and what did we learn?. *The Review of Corporate Finance Studies*, 4(2), 155-205.

Thakor, A. (2016). Corporate culture in banking.*Economic Policy Review*. The Federal Reserve Bank of New York, 5-16.

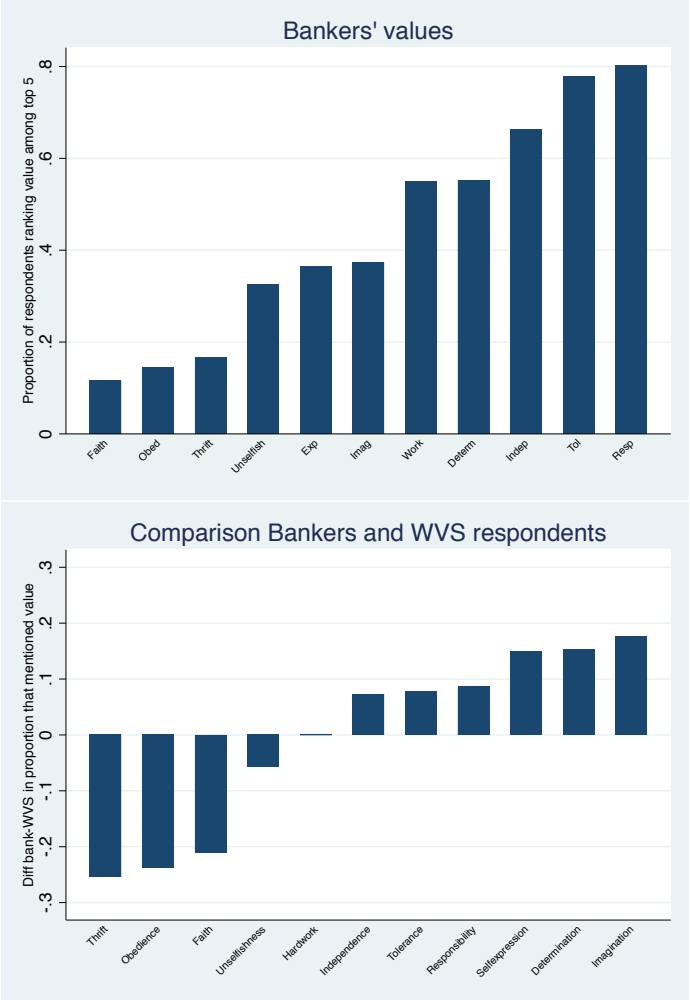
The Economist Intelligence Unit (2013). *A crisis of culture. Valuing ethics and knowledge in financial services*. A report from The Economist Intelligence Unit.

Van den Steen, E. (2010). On the origin of shared beliefs (and corporate culture). *The RAND Journal of Economics*, 41, 617-648.

Wendt, H., Euwema, M. & Van Emmerik, H. (2009). Leadership Styles and Team Cohesiveness Across Cultures. *The Leadership Quarterly*, 20, 358–370.

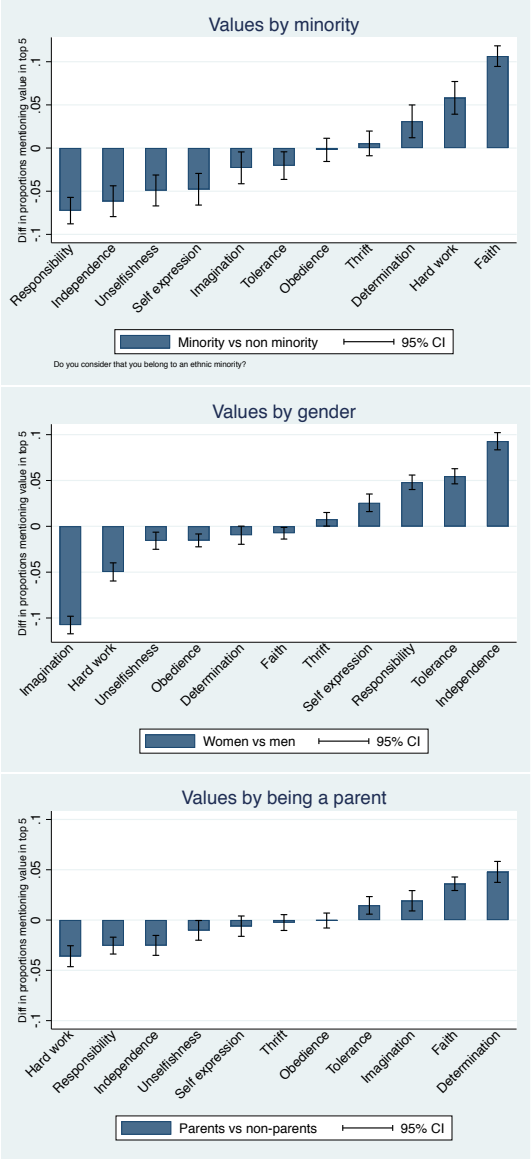
3.11 Figures

FIGURE 3.1: Bankers' values: overview and comparison with WVS



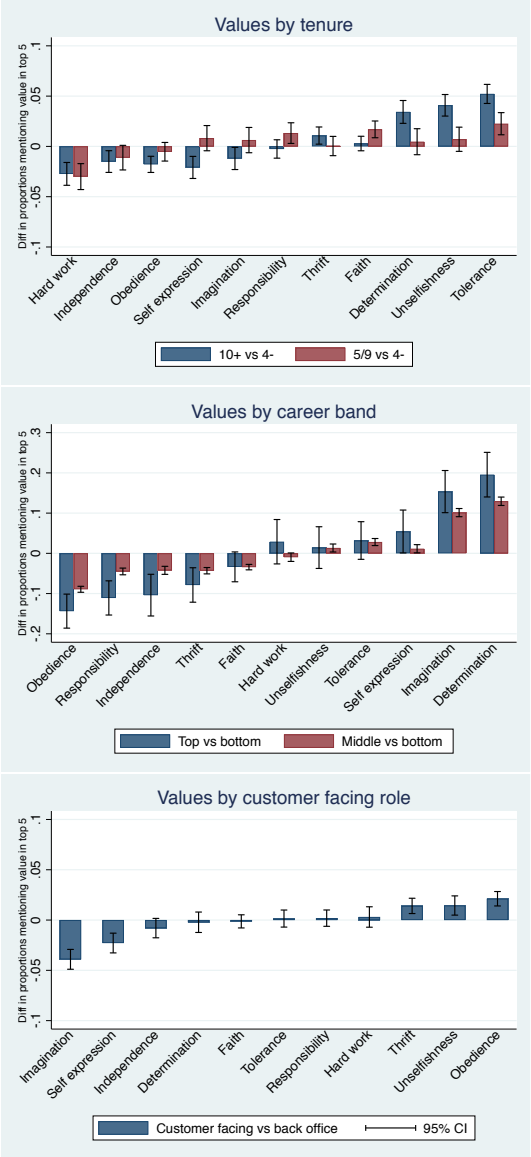
Note. The top figure shows the proportion of bankers who mentioned each of the given values among their top five. The bottom figure shows the percentage points difference in the proportion of people who mentioned a given value in the bank and in the WVS. Proportions are first computed country by country and averaged to get the illustrated overall difference.

FIGURE 3.2: Bankers' values by demographics



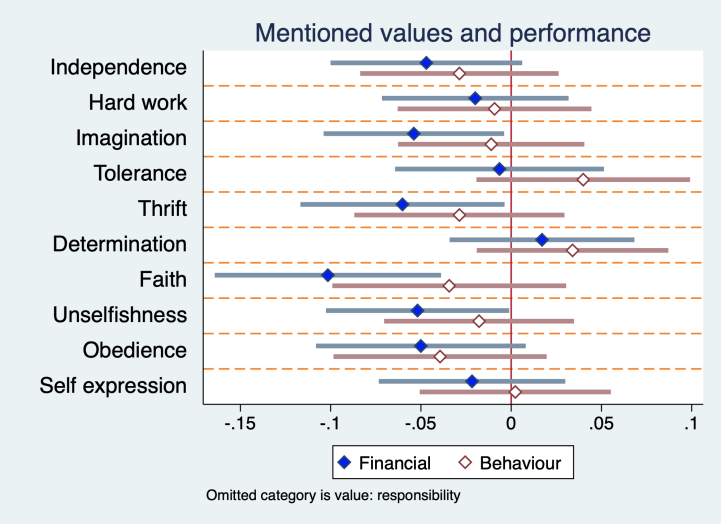
Note. The figures show differences in the proportions of people who mentioned a certain value by demographics. In the top figure, each bar gives the percentage points difference in the proportion of people who mentioned a given value between members and non-members of an ethnic minority. Members of a minority are defined as people who answer "yes" to the question: "Do you consider that you belong to an ethnic minority?". Most of people of Black, Latino or Mixed ethnicity answer "yes" to this question, while only 2% of White people and 30% of Asian people say "yes". The figure in the middle shows the percentage points difference in the proportion of people who mentioned a given value by gender. The bottom figure repeats the same exercise by parental status.

FIGURE 3.3: Bankers' values by organizational role



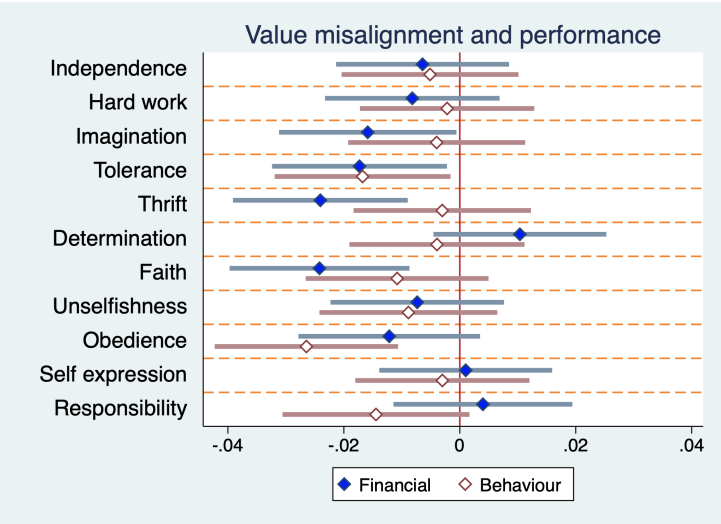
Note. The figures show differences in the proportions of people who mentioned a certain value by organizational role. In the top figure, each red bar gives the percentage points difference in the proportion of people who mentioned a given value between people with more than ten years of tenure and people with less than four years of tenure. Each blue bar gives the percentage points difference in the proportion of people who mentioned a given value between people with five to nine years of tenure and people with less than four years. The figure in the middle repeats this exercise by career band, classifying employees in three career bands. The bottom figure shows differences between employees in client facing roles and in back office.

FIGURE 3.4: Which values are correlated with performance?



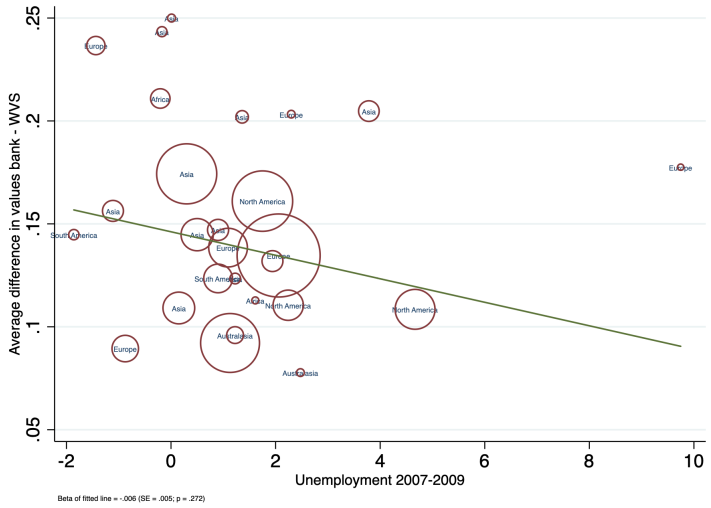
Note. The figure reports the coefficients of two separate OLS regressions of financial and behavioural performance on a set of ten indicator variables, one for each value that has been mentioned among the top five by a respondent. The omitted category is for the value "responsibility", which was mentioned by more than 85% of respondents. Horizontal bars show 90% confidence intervals.

FIGURE 3.5: Is misalignment in some values correlated with performance?



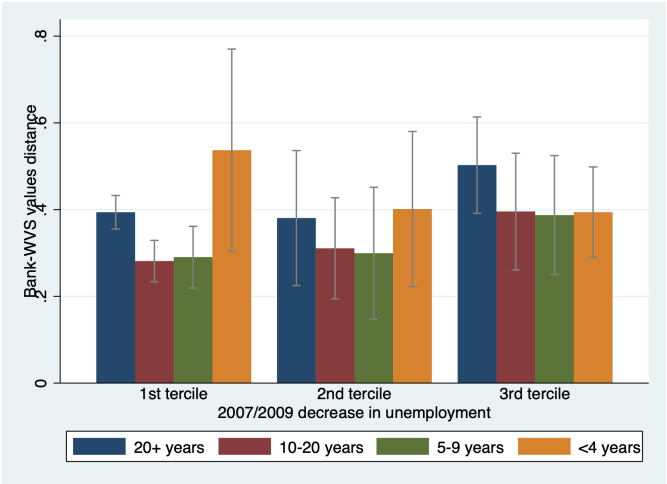
Note. The figure reports the coefficients of two separate OLS regressions of financial and behavioural performance on a set of eleven indexes of misalignment, one for each value. For instance, the coefficient on "imagination" is the marginal effect of an increase in the ranking distance of "imagination" between a respondent and his/her colleagues on the respondent's performance. Horizontal bars show 90% confidence intervals.

FIGURE 3.6: Bankers' distance with common citizens and the 2008 financial crisis



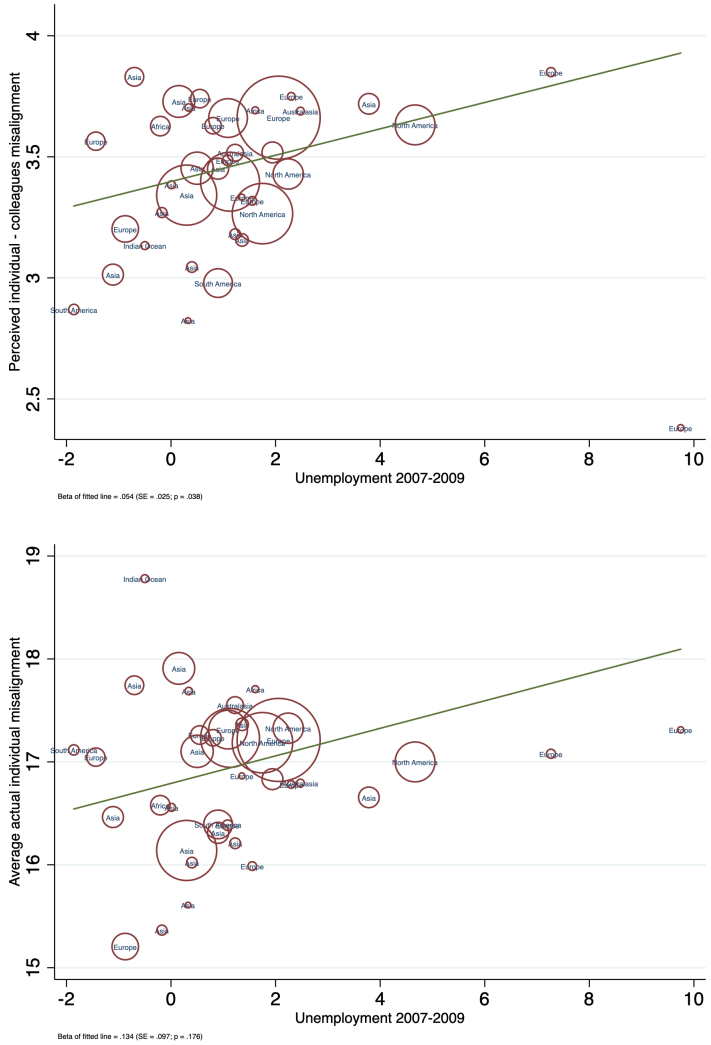
Note. Changes in unemployment, on the x-axis, are from Stevenson and Wolfers (2011). The y-axis reports the average distance between bankers and WVS respondents across the eleven values. The bubbles are proportional to the sample in each country and labelled with the country's region. The regression is estimated at the country level.

FIGURE 3.7: Bankers' distance with common citizens by tenure



Note. The bar chart shows average distance between bankers and WVS respondents by tenure and terciles of changes in unemployment rate during the 2008 financial crisis (from Stevenson and Wolfers, 2011).

FIGURE 3.8: Value misalignment and the 2008 financial crisis



Note. Changes in unemployment, on the x-axis, are from Stevenson and Wolfers (2011). The y-axis reports perceived misalignment with colleagues in the top scatterplot and actual misalignment with colleagues in the bottom figure. The bubbles are proportional to the sample in each country and labelled with the country's region. The regression is estimated at the country level.

3.12 Tables

TABLE 3.1: Summary statistics

		Values sample			Others			Diff
		Mean	SD	N	Mean	SD	N	p-val
Gender		.495	.5	38730	.501	.5	56143	.06
Asian		.692	.462	25350	.659	.474	32386	0
White		.274	.446	25350	.304	.46	32386	0
Black or Mixed		.035	.183	25350	.036	.187	32386	.23
Member of minority		.094	.292	30873	.096	.295	38939	.28
Age:	20-30	.254	.435	38750	.257	.437	56187	.33
	30-40	.388	.487	38750	.372	.483	56187	0
	40-50	.223	.416	38750	.226	.418	56187	.24
	50+	.135	.342	38750	.145	.352	56187	0
Tenure:	<1	.168	.374	38827	.17	.375	56311	.46
	1-4	.298	.457	38827	.301	.459	56311	.23
	5-9	.215	.411	38827	.208	.406	56311	.01
	10-20	.214	.41	38827	.208	.406	56311	.04
	20+	.106	.307	38827	.113	.316	56311	0
Customer facing role		.389	.487	38827	.387	.487	56311	.63
Assignment number		.143	.511	38107	.148	.526	55203	.22
Line manager		.293	.455	38827	.285	.452	56311	.01
Top career band		.008	.09	38827	.008	.087	56311	.46
Performance rating (std)		.005	.997	29037	-.004	1.002	41556	.26
Behaviour rating (std)		.001	1.002	27245	-.001	.999	38841	.85
Potential rating (std)		.005	1.003	17163	-.003	.998	25193	.44

Note. The table shows summary statistics for the sample of respondents to the questions on values ("Values sample") and for non-respondents ("Others"). "Assignment number" is equal to the total number of roles held by the employee in the bank. "Line manager" is an indicator variable for having line management responsibilities (self-reported). "Top career band" is an indicator variable for being in the top three hierarchical levels. "Performance", "Behaviour" and "Potential" ratings are the standardized measures of performance collected by HR.

TABLE 3.2: Value misalignment, individual and team performance

	DV: Performance score					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Financial		Behaviour		Potential	
	Panel A. Individual level					
Actual misalignment	-0.026***		-0.025***		-0.013	
	(0.010)		(0.009)		(0.014)	
Perceived misalignment		-0.024***		0.003		0.003
w/ colleagues		(0.005)		(0.006)		(0.014)
Perceived misalignment		-0.013		-0.026**		0.016
w/ bank		(0.011)		(0.010)		(0.014)
Observations	16,890	16,890	16,890	16,890	8,424	8,424
R-squared	0.058	0.059	0.033	0.033	0.064	0.064
	Panel B. Team level					
Actual misalignment	-0.006*		-0.007**		-0.004	
	(0.003)		(0.003)		(0.005)	
Perceived misalignment		-0.012		0.007		-0.007
w/ colleagues		(0.011)		(0.012)		(0.018)
Perceived misalignment		0.002		-0.028**		0.027*
w/ bank		(0.008)		(0.013)		(0.016)
Observations	8,741	8,741	8,741	8,741	5,442	5,442
R-squared	0.055	0.054	0.051	0.051	0.067	0.067
Controls	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y

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

Note. OLS regressions. The dependent variable is the standardized measure of financial performance in Columns (1) and (2), behavioural performance in Columns (3) and (4) and potential for progression in Columns (5) and (6). Panel A uses data at the individual level. Actual misalignment for individuals is defined as the Absolute Spearman ρ index of misalignment with their team colleagues and manager. Perceived misalignment is the distance (in terms of Absolute Spearman ρ) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Panel B aggregates data at the team level. Actual and perceived team misalignment are obtained by averaging across individuals (without double-counting pairs of employees). Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.3: Value misalignment with manager and individual performance

	DV: Individual performance score								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Financial			Behaviour			Potential		
Actual misgnment	-0.027*			-0.030*			0.030		
w/ coll + mng	(0.016)			(0.017)			(0.026)		
Actual misgnment		-0.025**	-0.023**		-0.028*	-0.026*		0.009	0.007
w/ manager		(0.010)	(0.010)		(0.015)	(0.015)		(0.021)	(0.021)
Actual misgnment		-0.010	-0.007		-0.006	-0.003		0.019	0.016
w/ colleagues		(0.016)	(0.016)		(0.013)	(0.013)		(0.023)	(0.024)
Perceived misgnment			-0.043***						
w/ colleagues			(0.012)						
Perceived misgnment						-0.030			0.034
w/ bank						(0.018)			(0.030)
Observations	4,428	4,428	4,428	4,428	4,428	4,428	2,084	2,084	2,084
R-squared	0.076	0.076	0.078	0.040	0.040	0.041	0.098	0.098	0.099
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Mean DV	0.03	0.03	0.03	0	0	0	-0.03	-0.03	-0.03

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

Note. OLS regressions. The dependent variable is the standardized measure of financial performance in Columns (1) to (3), behavioural performance in Columns (4) to (6) and potential for progression in Columns (7) to (9). The table uses data at the individual level. The sample size is determined by the number of people whose manager is also a respondent to the values questions. Actual misalignment for individuals is defined as the Absolute Spearman ρ index of misalignment with their team colleagues and manager in the first row, with the manager only in the second row and with colleagues only in the third row. Perceived misalignment is the distance (in terms of Absolute Spearman ρ) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.4: Value misalignment and trust

VARIABLES	(1) Trust in coll	(2) Trust in mng	(3) Trust leaders	(4) Feel trusted
Actual misgnment	-0.014*	-0.000	-0.009**	-0.017***
w/ coll + mng	(0.007)	(0.004)	(0.004)	(0.005)
Perceived misgnment	-0.045***	-0.046***	-0.015	-0.020***
w/ colleagues	(0.009)	(0.006)	(0.012)	(0.007)
Perceived misgnment	-0.060***	-0.053***	-0.107***	-0.079***
w/ bank	(0.014)	(0.012)	(0.022)	(0.016)
Observations	23,657	23,657	23,657	23,657
R-squared	0.045	0.048	0.131	0.082
Controls	Y	Y	Y	Y
Country FE	Y	Y	Y	Y

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

Note. OLS regressions. The dependent variables are different measures for trust in others. In Columns (1), (2) and (3) the dependent variable is a self-reported measure of trust in colleagues, the manager and senior leaders. The variable is measured on a scale from 1 (min) to 5 (max). The dependent variable in Column (4) is a measure of the extent to which the respondent agrees with the statement "Senior leadership in my function trust me to do what's right" on a scale from 1 (min) to 5 (max). The table uses data at the individual level. Actual misalignment for individuals is defined as the Absolute Spearman ρ index of misalignment with their team colleagues and manager. Perceived misalignment is the distance (in terms of Absolute Spearman ρ) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.5: Managers' attitudes and team culture

DV: Managers' attitudes and team misalignment with bank							
VARIABLES	(1) Not enough time	(2) Not enough product	(3) Not proud of work	(4) Not confident in future	(5) Disagree w/ leaders	(6) Not confident in colleagues	(7) Misaligned w/ bank
Perceived misgnment w/ colleagues	0.005	0.001	0.000	0.000	0.004	0.004	-0.002
	(0.009)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.012)
Perceived misgnment w/ bank	0.009	0.001	0.008*	0.014**	0.010*	-0.005	0.025**
	(0.007)	(0.009)	(0.004)	(0.006)	(0.006)	(0.003)	(0.012)
Observations	5,282	5,282	5,282	5,282	5,282	5,282	1,734
R-squared	0.111	0.157	0.120	0.117	0.126	0.064	0.155
Actual misgnment w/ colleagues	-0.001	-0.002	-0.000	0.000	0.001	-0.001	-0.002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Observations	5,363	5,363	5,363	5,363	5,363	5,363	1,756
R-squared	0.113	0.159	0.118	0.114	0.122	0.063	0.150
Controls	Y	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y	Y

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

Note. OLS regressions. The dependent variables are measures of managers' attitudes towards their work and the bank. In Columns (1) to (3) the dependent variables are managers' opinion of their productivity: whether they think there is not enough time to do things properly, whether they feel that conditions in the job do not allow them to be as productive as possible and whether they are proud of their work. In Columns (4) to (6) the dependent variables are managers' confidence in the future, disagreement with leaders and confidence in colleagues. In Column (7) the dependent variable is managers' own perceived misalignment in values with the bank. Original variables are coded in a five steps Likert scale. The dependent variables are indicator variables for the two highest scores in the scales. Independent variables are at the team level (each team has only one manager). Actual misalignment for individuals is defined as the Absolute Spearman ρ index of misalignment with their team colleagues and manager. Perceived misalignment is the distance (in terms of Absolute Spearman ρ) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Actual and perceived team misalignment are obtained by averaging across individuals (without double-counting pairs of employees). Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.6: Misalignment and team communication/coordination

VARIABLES	(1) Communication Index	(2) Coordination Index
Actual misalignment	-0.006	0.004
w/ coll + mng	(0.008)	(0.004)
Perceived misalignment	-0.026***	-0.015***
w/ collagues	(0.006)	(0.004)
Perceived misalignment	-0.038***	-0.007
w/ bank	(0.005)	(0.005)
Observations	9,107	9,107
R-squared	0.151	0.333
Controls	Y	Y
Country FE	Y	Y

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

Note. OLS regressions. The communication index is the average of answers to the following questions: "How frequently do you have team meetings?", "How often do you check-in with your manager to discuss your development?", "Have you participated in informal team meetings in the last 3 months?", "Do you discuss how your work contributes to the strategy of the bank with the manager?" "Does your manager give you feedback to improve?". The coordination index is the number of times that a person agrees with the following statements: "I feel able to speak up when I see behaviour which I consider to be wrong","Where I work, issues of right and wrong are discussed in team meetings", "Where I work, colleagues take responsibility for their actions", "Where I work, everyone has the chance to say what they think about issues", "Where I work, people can state their opinion without the fear of negative consequences". Both indexes are constructed at the individual level and then averaged at the team level. Actual misalignment for individuals is defined as the Absolute Spearman ρ index of misalignment with their team colleagues and manager. Perceived misalignment is the distance (in terms of Absolute Spearman ρ) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Actual and perceived team misalignment are obtained by averaging across individuals (without double-counting pairs of employees). Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.7: Introducing extrovert and obedient employees

	Extrovert (1)		Obedient (2)		
	Margin	SE	Margin	SE	Diff (2) - (1)
Imagination	0.501	0.008	0.209	0.009	-0.292
Self expression	0.477	0.008	0.208	0.009	-0.269
Independence	0.762	0.005	0.587	0.008	-0.175
Determination	0.607	0.006	0.507	0.008	-0.100
Tolerance	0.806	0.005	0.740	0.006	-0.066
Responsibility	0.843	0.004	0.834	0.005	-0.010
Unselfishness	0.315	0.005	0.365	0.007	0.049
Hard work	0.506	0.007	0.645	0.008	0.140
Thrift	0.089	0.006	0.292	0.007	0.204
Faith	0.031	0.004	0.241	0.008	0.210
Obedience	0.029	0.006	0.332	0.012	0.303

Note. The table shows the proportion of people that mentioned each of the eleven values in the two classes identified by latent class analysis. We label employees in class 1 as "extrovert" and in class 2 as "obedient".

TABLE 3.8: Individual performance: extrovert vs obedient types

DV: Performance score			
	(1) Financial	(2) Behaviour	(3) Potential
Extrovert	0.042*	0.029***	0.093***
	(0.022)	(0.010)	(0.020)
Observations	16,890	16,890	8,424
R-squared	0.058	0.032	0.066
Mean DV	0.02	0.01	0.01

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

Note. OLS regressions. The dependent variable is the standardized measure of financial performance in Column (1) , behavioural performance in Column (2) and potential for progression in Column (3). The table uses data at the individual level. The sample size is determined by the number of people whose manager is also a respondent to the values questions. The variable "Extrovert" is an indicator equal to one if the person was categorized by the latent class analysis as an extrovert" type. Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.9: Extroverts' share and team performance

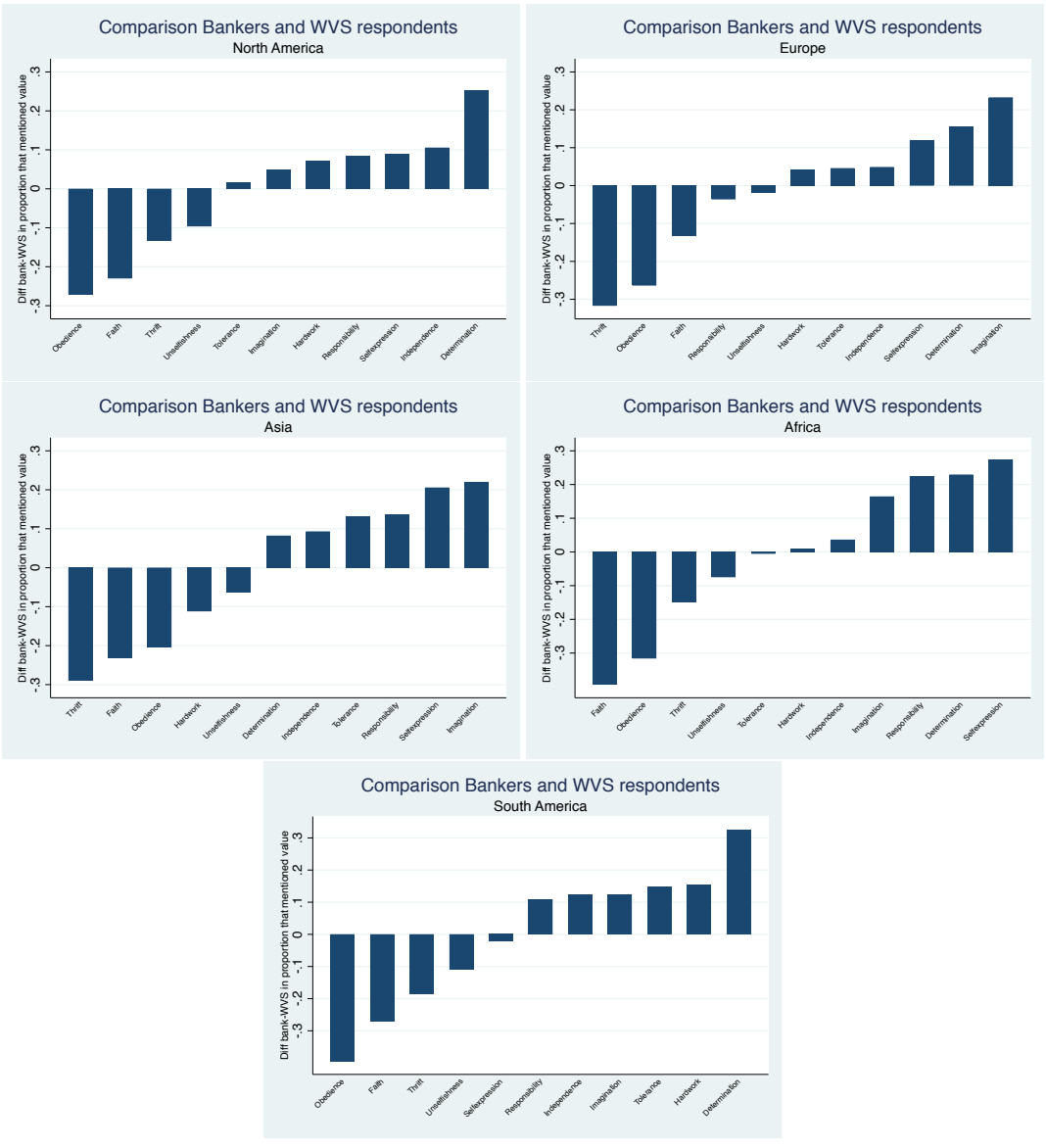
DV: Performance score						
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Financial		Behaviour		Potential	
Extrovert prop $\notin [0.3, 0.6]$	0.044**		0.045		0.061**	
	(0.021)		(0.042)		(0.027)	
Extrovert prop $\in [0, 0.3)$		0.005		-0.010		0.006
		(0.026)		(0.034)		(0.045)
Extrovert prop $\in (0.6, 1]$		0.058*		0.064		0.076**
		(0.029)		(0.048)		(0.030)
Observations	8,871	8,871	8,871	8,871	5,548	5,548
R-squared	0.053	0.053	0.050	0.050	0.067	0.068
Controls	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y
Mean DV	0.0100	0.0100	0	0	0	0

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

Note. OLS regressions. The dependent variable is the standardized measure of financial performance in Columns (1) and (2), behavioural performance in Columns (3) and (4) and potential for progression in Columns (5) and (6). The variable "Extrovert prop $\notin [0.3, 0.6]$" is equal to one if a team has a share of extrovert employees below 30% or above 60%. Similarly, the variables "Extrovert prop $\in (0.6, 1]$" and "Extrovert prop $\in [0, 0.3)$" are equal to one if a team has a share of extrovert employees above 60% or below 30% respectively. Data are at the team level. Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

3.A Appendix figures and tables

FIGURE 3.A.1: Bankers vs WVS respondents by continent



Note. The figures show the percentage points difference in the proportion of people who mentioned a given value in the bank and in the WVS, by continent. Proportions are first computed country by country and averaged to get the illustrated overall difference in each continent.

TABLE 3.A.1: Which values are correlated with performance?

DV: Performance score			
	(1)	(2)	(3)
VARIABLES	Financial	Behaviour	Potential
Independence	-0.047	-0.029	0.001
	(0.033)	(0.032)	(0.023)
Hard work	-0.020	-0.009	0.034
	(0.028)	(0.024)	(0.028)
Imagination	-0.054*	-0.011	0.111***
	(0.028)	(0.031)	(0.025)
Tolerance	-0.006	0.040	0.024
	(0.033)	(0.040)	(0.039)
Thrift	-0.060***	-0.029	0.000
	(0.020)	(0.022)	(0.032)
Determination	0.017	0.034	0.128***
	(0.024)	(0.025)	(0.025)
Faith	-0.102**	-0.034	-0.018
	(0.047)	(0.055)	(0.051)
Unselfishness	-0.052*	-0.018	0.035
	(0.027)	(0.029)	(0.054)
Obedience	-0.050	-0.039	-0.029
	(0.035)	(0.041)	(0.033)
Self expression	-0.022	0.002	0.081**
	(0.031)	(0.032)	(0.031)
Observations	16,890	16,890	8,424
R-squared	0.020	0.018	0.033
Controls	Y	Y	Y
Country FE	Y	Y	Y
Mean DV	0.02	0.01	0.01

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

Note. OLS regressions of financial, behavioural and potential performance (standardized) on a set of ten indicator variables, one for each value that has been mentioned among the top five by a respondent. The omitted category is for the value “responsibility", which was mentioned by more than 85% of respondents.Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.A.2: Is misalignment in some values correlated with performance?

DV: Performance score			
	(1)	(2)	(3)
VARIABLES	Financial	Behaviour	Potential
Independence	-0.006	-0.005	-0.008
	(0.010)	(0.006)	(0.009)
Hard work	-0.008	-0.002	-0.009
	(0.006)	(0.008)	(0.018)
Imagination	-0.016	-0.004	0.022***
	(0.012)	(0.009)	(0.007)
Tolerance	-0.017**	-0.017**	-0.006
	(0.008)	(0.008)	(0.013)
Thrift	-0.024***	-0.003	-0.014
	(0.009)	(0.007)	(0.011)
Determination	0.010	-0.004	-0.003
	(0.011)	(0.009)	(0.011)
Faith	-0.024***	-0.011	-0.020*
	(0.009)	(0.009)	(0.012)
Unselfishness	-0.007	-0.009	0.002
	(0.010)	(0.007)	(0.014)
Obedience	-0.012	-0.026***	-0.029***
	(0.008)	(0.010)	(0.011)
Self expression	0.001	-0.003	0.003
	(0.005)	(0.005)	(0.009)
Responsibility	0.004	-0.014	0.008
	(0.009)	(0.010)	(0.012)
Observations	16,890	16,890	8,424
R-squared	0.020	0.017	0.027
Controls	Y	Y	Y
Country FE	Y	Y	Y
Mean DV	0.02	0.01	0.01

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

Note. OLS regressions of financial, behavioural and potential performance (standardized) on a set of eleven indexes of misalignment, one for each value. For instance, the coefficient on "imagination" is the marginal effect of an increase in the ranking distance of "imagination" between a respondent and his/her colleagues on the respondent's performance (measures in terms of Absolute Spearman ρ). Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.A.3: Summary statistics by extrovert and obedient type

	Obedient			Extrovert			Diff
	Mean	SD	N	Mean	SD	N	p-val
Gender	.479	.5	16931	.507	.5	21799	0
Asian	.696	.46	11250	.688	.463	14100	.16
White	.262	.44	11250	.282	.45	14100	0
Black or Mixed	.041	.198	11250	.029	.169	14100	0
Member of minority	.112	.316	13457	.08	.271	17416	0
Age 20-30	.253	.435	16940	.255	.436	21810	.62
Age 30-40	.386	.487	16940	.389	.488	21810	.46
Age 40-50	.221	.415	16940	.225	.417	21810	.33
Age 50+	.141	.348	16940	.131	.337	21810	0
Tenure <1	.172	.377	16984	.164	.371	21843	.05
Tenure 1-4	.295	.456	16984	.3	.458	21843	.22
Tenure 5-9	.212	.409	16984	.217	.412	21843	.29
Tenure 10-20	.209	.407	16984	.218	.413	21843	.04
Tenure 20+	.112	.315	16984	.101	.301	21843	0
Customer facing	.402	.49	16984	.379	.485	21843	0
Assignment number	.131	.477	16821	.153	.537	21286	0
Line manager	.287	.452	16984	.298	.457	21843	.02
Top career band	.006	.078	16984	.01	.098	21843	0
Performance rating (std)	-.037	.995	12786	.038	.998	16251	0
Values rating (std)	-.036	1.007	12158	.03	.996	15087	0
Potential rating (std)	-.041	1.005	7178	.037	1	9985	0

Note. The Table reports summary statistics by extrovert and obedient types. “Assignment number" is equal to the total number of roles held by the employee in the bank. “Line manager" is an indicator variable for having line management responsibilities (self-reported). “Top career band" is an indicator variable for being in the top three hierarchical levels. “Performance", “Behaviour" and “Potential" ratings are the standardized measures of performance collected by HR.

TABLE 3.A.4: Individual value misalignment and team coordination

VARIABLES	(1) Frequent meetings w/ team	(2) w/ mngr	(3) Informal meetings	(4) Talked about strategy	(5) feedback	(6) Coordination index
Actual misgnment	-0.006	0.027*	-0.003	-0.003	0.000	0.003
	(0.007)	(0.014)	(0.003)	(0.004)	(0.003)	(0.003)
Perceived misgnment w/ collagues	-0.036***	-0.033***	-0.010*	-0.008***	-0.019***	-0.021***
	(0.007)	(0.008)	(0.005)	(0.002)	(0.003)	(0.002)
Perceived misgnment w/ bank	-0.023**	-0.036***	-0.005	-0.016***	-0.016***	-0.019***
	(0.009)	(0.009)	(0.004)	(0.004)	(0.002)	(0.004)
Observations	23,657	23,657	23,657	23,657	23,657	23,657
R-squared	0.121	0.174	0.045	0.050	0.044	0.178
Controls	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y

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

Note. OLS regressions of each component of the communication index on actual and perceived misalignment at the individual level. Variables in each of the first five Columns are constructed from answers to the following questions: "How frequently do you have team meetings?", "How often do you check-in with your manager to discuss your development?", "Have you participated in informal team meetings in the last 3 months?", "Do you discuss how your work contributes to the strategy of the bank with the manager?" "Does your manager give you feedback to improve?". Column (6) averages the answers to these questions to construct an index. Actual misalignment for individuals is defined as the Absolute Spearman ρ index of misalignment with their team colleagues and manager. Perceived misalignment is the distance (in terms of Absolute Spearman ρ) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

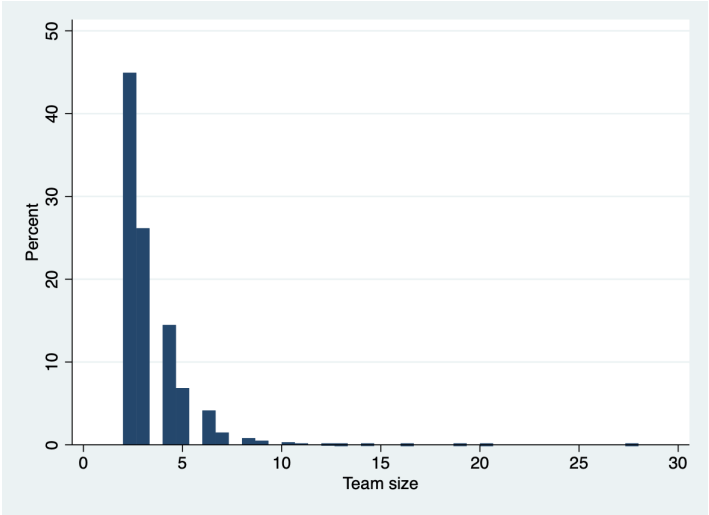
TABLE 3.A.5: Individual value misalignment and team communication

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Speak up		Colleagues take responsibility	Freedom of speak up	opinion	Communication index
Actual misgnment	-0.011**	-0.002	-0.008*	-0.015**	-0.003	-0.008
	(0.004)	(0.007)	(0.005)	(0.007)	(0.009)	(0.006)
Perceived misgnment w/ collagues	-0.051***	-0.045***	-0.063***	-0.050***	-0.062***	-0.054***
	(0.008)	(0.007)	(0.011)	(0.010)	(0.009)	(0.008)
Perceived misgnment w/ bank	-0.055***	-0.073***	-0.039***	-0.053***	-0.084***	-0.061***
	(0.008)	(0.012)	(0.008)	(0.012)	(0.012)	(0.010)
Observations	23,657	23,657	23,657	23,657	23,657	23,657
R-squared	0.044	0.069	0.055	0.039	0.056	0.067
Controls	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y

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

Note. OLS regressions of each component of the communication index on actual and perceived misalignment at the individual level. Variables in each of the first five Columns are constructed from agreement with the following statements:"I feel able to speak up when I see behaviour which I consider to be wrong","Where I work, issues of right and wrong are discussed in team meetings", "Where I work, colleagues take responsibility for their actions", "Where I work, everyone has the chance to say what they think about issues", "Where I work, people can state their opinion without the fear of negative consequences". Column (6) averages the answers to these questions to construct an index. Actual misalignment for individuals is defined as the Absolute Spearman ρ index of misalignment with their team colleagues and manager. Perceived misalignment is the distance (in terms of Absolute Spearman ρ) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

FIGURE 3.A.2: Team size distribution



Note. The figure shows the distribution of team sizes in the sample.

TABLE 3.A.6: Kendall τ index: value misalignment, individual and team performance

	DV: Performance score					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Financial		Behaviour		Potential	
	Panel A. Individual level					
Actual misalignment	-0.021***		-0.031***		-0.010	
	(0.007)		(0.010)		(0.010)	
Perceived misalignment		-0.024***		0.003		0.003
w/ colleagues		(0.005)		(0.006)		(0.014)
Perceived misalignment		-0.013		-0.026**		0.016
w/ bank		(0.011)		(0.011)		(0.014)
Observations	16,890	16,890	16,890	16,890	8,424	8,424
R-squared	0.058	0.059	0.033	0.033	0.064	0.064
	Panel B. Team level					
Actual misalignment	-0.006*		-0.007**		-0.004	
	(0.004)		(0.003)		(0.005)	
Perceived misalignment		-0.003		0.001		-0.002
w/ colleagues		(0.002)		(0.003)		(0.004)
Perceived misalignment		0.000		-0.006**		0.006
w/ bank		(0.002)		(0.003)		(0.004)
Observations	8,741	8,741	8,741	8,741	5,442	5,442
R-squared	0.055	0.054	0.051	0.051	0.067	0.067
Controls	Y	Y	Y	Y	Y	Y
Country FE	Y	Y	Y	Y	Y	Y

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

Note. OLS regressions. The dependent variable is the standardized measure of financial performance in Columns (1) and (2), behavioural performance in Columns (3) and (4) and potential for progression in Columns (5) and (6). Panel A uses data at the individual level. Actual misalignment for individuals is defined as the Kendall τ index of misalignment with their team colleagues and manager. Perceived misalignment is the distance (in terms of Kendall τ) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Panel B aggregates data at the team level. Actual and perceived team misalignment are obtained by averaging across individuals (without double-counting pairs of employees). Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.

TABLE 3.A.7: Kendall τ index: value misalignment and trust

VARIABLES	(1) Trust in coll	(2) Trust in mng	(3) Trust leaders	(4) Feel trusted
Actual misgnment	-0.016***	-0.005	-0.012**	-0.015**
	(0.006)	(0.007)	(0.006)	(0.006)
Perceived misgnment w/ bank	-0.061***	-0.054***	-0.109***	-0.081***
	(0.014)	(0.012)	(0.023)	(0.017)
Perceived misgnment w/ colleagues	-0.046***	-0.047***	-0.015	-0.020***
	(0.009)	(0.006)	(0.012)	(0.007)
Observations	23,657	23,657	23,657	23,657
R-squared	0.046	0.049	0.131	0.082

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

Note. OLS regressions. The dependent variables are different measures for trust in others. In Columns (1), (2) and (3) the dependent variable is a self-reported measure of trust in colleagues, the manager and senior leaders. The variable is measured on a scale from 1 (min) to 5 (max). The dependent variable in Column (4) is a measure of the extent to which the respondent agrees with the statement "Senior leadership in my function trust me to do what's right" on a scale from 1 (min) to 5 (max). The table uses data at the individual level. Actual misalignment for individuals is defined as the Kendall τ index of misalignment with their team colleagues and manager. Perceived misalignment is the distance (in terms of Kendall τ index) between the values that respondents choose for themselves and the ones they choose for their colleagues or the bank. Controls include job role, gender, tenure, customer facing role, ethnicity, top career band and assignment number.