ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Gaudeul, Alexia; Okvitawanli, Ayu; Panganiban, Marian

Working Paper

Does the gender mix among employers influence who gets hired? A labor market experiment

Jena Economic Research Papers, No. 2015-007

Provided in Cooperation with:
Friedrich Schiller University Jena, Faculty of Economics and Business Administration

Suggested Citation: Gaudeul, Alexia; Okvitawanli, Ayu; Panganiban, Marian (2015) : Does the gender mix among employers influence who gets hired? A labor market experiment, Jena Economic Research Papers, No. 2015-007, Friedrich Schiller University Jena, Jena

This Version is available at:
https://hdl.handle.net/10419/121281

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

JENA ECONOMIC RESEARCH PAPERS

2015 – 007

Does the gender mix among employers influence who gets hired? A labor market experiment

by

Alexia Gaudeul
Ayu Okvitawanli
Marian Panganiban

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

© by the author.

Does the gender mix among employers influence who gets hired? A labor market experiment☆

Alexia Gaudeul[a], Ayu Okvitawanli[a], and Marian Panganiban[a,b,*]

[a]*Friedrich-Schiller-Universität, Jena, Germany*
[b]*Max Planck Institute for Research on Collective Goods, Bonn, Germany*

Abstract

We consider in this paper whether the gender mix at the level of decision-makers in firms can influence gender representation at the employee level. We run a laboratory experiment whereby we present a pair of independent employers with applications from two potential employees. We consider whether the gender of the other employer will influence an employer's hiring decision. We find that the gender mix among employers plays a role in the individual hiring decisions of female members. Female employers when paired with a male employer are more likely to choose a female applicant over an equally competent male applicant. Results of an Implicit Association Test (IAT) and answers to a post-experimental questionnaire show that explicit beliefs about relative gender performance are significantly associated with the observed hiring bias, while implicit attitudes do not appear to play a role.

Keywords: discrimination, hiring, IAT, implicit attitudes, gender quotas, labor markets, employment
JEL-codes: J71, J78, C91

☆We thank Oliver Kirchkamp, Olexandr Nikolaychuk, and Marie Claire Villeval for their help and comments, as well as participants in seminars held at the Max Planck Institute of Economics and at the Friedrich Schiller University in Jena, Germany in 2014. We acknowledge the financial support of the Max Planck Institute of Economics and the excellent assistance of Friedrich Gehring and Claudia Zellman in conducting the experiment. We also wish to acknowledge the hospitality of McDonald's restaurant in Jena and its conducive research atmosphere.

*Corresponding Author. e-mail: panganiban@econ.mpg.de

1. Introduction

Claudia Goldin (2014) writes in her presidential address to the American Economic Association that one of the grandest advances of the last century has been "the converging roles of men and women in society and the economy." The last fifty years have indeed seen significant progress in the educational attainment, political representation, and labor force participation of women (Pande and Ford (2012); Azmat and Petrongolo (2014)). The United Nations Millennium Development Goals Report states that all developing regions have achieved, or were close to achieving, gender parity in primary education in 2012 (Millenium Development Report, 2014). By 1994, women had obtained the right to vote in 96% of the countries in the world (Ramirez et al., 1997). Growth in the labor force was higher for women than for men in every region in the world except Africa in the 1980s and early 1990s (Lim, 2009). However, these overall improvements in female labor force and political participation have not yet translated into a corresponding increase of women in leadership positions. Less than 24% of legislators in the Parliament around the world are female (Women in Politics, IPU and UN, 2015). In the corporate sector, female representation also declines with higher positions. To illustrate, only 17.8% of board members of the largest publicly listed companies in the European Union are women (EuropeanCommission, 2014), in contrast to a female labor force participation rate of 65.8% (World Bank, 2014).

One direct but still widely debated policy to address the marginal representation of women in leadership positions is the introduction of gender quotas. The central idea behind quotas is that the proportion of women in decision-making bodies should not be lower than a certain level. In political representation, gender quotas can come in the form of having seats reserved for women in the parliament, imposing a minimum number of females in the candidate lists, or as measures written in the statutes of parties (Women in Politics, IPU and UN, 2015).

In the corporate world, quotas come in the form of legislated ratios of female representation in the corporate board and / or senior management. Some countries that have enacted corporate board quotas include Norway, France, Belgium, and Canada.

Our paper investigates a key assumption of the gender quota policy, namely that changes in the gender composition of a decision-making body can influence the individual decisions of its members. We investigate this in the context of hiring employees. Specifically, we consider whether the gender mix at the level of employers affects individual hiring decisions. We conducted a laboratory experiment where we matched two employers with two applicants, gave information about the level of competence, gender, age and education of the applicants to the employers, and observed consequences of changing the gender mix among employers on the number of males and females hired. We structured the pay-offs of employers and applicants to eliminate strategic and other-regarding concerns so as to focus on the effects of taste-based discrimination.

The next few paragraphs review related literature on gender quotas and hiring discrimination and outline our main contribution.

1.1. Gender quotas

A motivation for our work is an assumption implicit in the imposition of gender quotas in leadership position, which is that higher representation of women at senior levels will improve the position of women at lower levels and foster their advancement. Previous studies on gender quotas have mostly investigated its impact in terms of equity and efficiency goals such as achieving wage equality or improving company performance. Given this research focus, most research uses large data sets at the country- (Ahern and Dittmar, 2012) or at the firm-level (Chambliss and Uggen, 2000; Gorman, 2005). One challenge in this research is that it is difficult to make causal claims in terms of the impact of gender quotas on efficiency and equity (Pande and Ford, 2012). Indeed, gender quotas are often legislated at the same time as

the notion of equality in leadership and representation becomes more widely accepted (Krook, 2006). Changes in outcomes can thus be attributed to changing attitudes towards female representation rather than to a quota policy *per se*. Another issue is that most gender quotas have only been implemented recently, and their effect on efficiency or equity might take a long time to be established. Our idea in this experiment is to consider how gender quotas affect individual decision-making. Indeed, underlying the discussion on gender quota's relationship with equitable or efficient outcomes is its impact on individual preferences and behavior. In this case, instead of using firm or country data, conducting experiments that allow us to investigate individual decisions might not only be more feasible, but also useful.

We focus on hiring because it is an observable outcome that can be directly linked with the gender quota policy to the extent that corporate board members or senior management are directly involved in hiring decisions. While discrimination in hiring has been investigated in many empirical and theoretical studies, few have related hiring discrimination to the gender composition of the pool of employers. The closest studies to ours in this regard are by Bagues and Esteve-Volart (2010), which uses a repeated randomized experiment to test if the gender composition of recruiting committees affects the chances of success of 150 000 female and male candidates for positions in the Spanish Judiciary, and by Zinovyeva and Bagues (2010) which also uses a randomized experiment to see how the proportion of female evaluators increases the chances of success of female applicants to full professor positions in Spain. Both studies find some evidence of a positive relationship between the proportion of female members in the pool of employers with the proportion of women being hired.

Following this line of research, we conduct a laboratory experiment with university students as participants to investigate discrimination at the individual level. We believe that the use of university students as participants is appropriate in this context given that they are expected to pursue careers in the corporate or policy sector and assume

leadership possitions. The laboratory environment allows us to control the composition of the pool of employers, the pairs of applicants that are presented to them, the task that applicants are asked to perform, and the information both parties have about each other. Thus, we avoid the identification challenge in some empirical studies that are caused by the distribution of applicants being skewed towards one gender, or the possibility that requirements of the position favor one gender over the other. In this regard, our work is related to Bendick Jr. et al. (1994) who use experiments to uncover racial discrimination in hiring and to Balafoutas and Sutter (2012) who test in the laboratory the effects of different policy interventions, one of which is quotas, on the likelihood that women take part in competitions.

1.2. Statistical and taste-based discrimination

While we wish to investigate the role of the gender composition of the hiring pool in hiring decisions, we need to control for two main possible drivers of discrimination, statistical and taste-based discrimination. Previous studies have documented those drivers of discrimination in the labor market (see Darity and Mason (1998) for a comprehensive summary). Statistical discrimination (Arrow, 1973; Aigner and Cain, 1977) occurs as a result of imperfect information about the performance of potential candidates. One's social group, in this case gender, can be used as an indicator of performance instead of actual performance. To illustrate, male employers in a male-dominated occupation will favor males because they have relatively more information about male performance and thus perceive the choice of females as more risky. This risk aspect can drive the selection of a male applicant even against a female candidate that appears to be more competent. As males employers disproportionally select males, they continue to receive more information about male performance and less information about female performance, thus contributing to the persistence of a bias against hiring females. Male and female candidates then have different incentives to invest in human capital as the probability to get returns on their in-

vestments in a particular occupation depend on their gender (see also Beaman et al. (2009)). Some female applicants may then opt not to apply for male-dominated jobs, and those already employed in those jobs have less incentives to invest the time and effort to advance their position in the company. This mechanism may be one of the reasons that males maintain their representation in senior positions (Burgess and Tharenou, 2002).

Gender discrimination in the labor market can manifest itself not only because of differences in the quality and quantity of information employers receive on applicants, but also because employers have a preference for one gender (usually male) over another. This is what Becker (1971) modelled as taste-based discrimination. Research in psychology has long investigated the many ways this preference manifests itself. In the workplace, for example, females are usually perceived as less competent and less productive than males (Huddy and Terkildsen, 1993). People in hiring positions positions act upon this stereotype even when presented with females that are more competent and more productive than males. In a study by Steinpreis et al. (1999), identical scientific resumes were sent to 238 male and female academics. These resumes either have a male-sounding or female-sounding applicants' name. The academics were asked if they would accept the applicants as their working colleagues. Female and male academics accepted significantly more male applicants than female applicants. Moss-Racusin et al. (2012) also find similar results in their study. Moreover, they find that perceived competence mediates hiring decision. That is, females are perceived as less competent and consequently, hired less often. They also found that females with an identical resume were hired with lower a starting salary and with a lower mentoring commitment.

Models of statistical and taste-based discrimination do not necessarily mean that individuals discriminate consciously. Indeed, Bertrand et al. (2005) focus rather on implicit attitudes as a driver of discrimination. In our experiment therefore, we not only controlled for statistical discrimination by asking participants whether they perceive males as

higher performing than women in the task applicants were asked to perform for them, but we checked for the implicit attitudes of our participants. We implemented an Implicit Association Test ("IAT") similar to what Rooth (2010) did to analyze hiring discrimination among applicants with Arab-sounding or Swedish-sounding names. We also had participants answer a post-experimental questionnaire to check their explicit atittudes on gender to take into account possible discrepancies between overt expressions of hiring prejudice and implicit attitudes (Rudman and Kilianski, 2000).

The paper proceeds by presenting our hypothesis in Section 2. Section 3 describes the experimental design. Section 4 presents and discusses the results while Section 5 concludes.

2. General hypotheses

We begin with a model of Becker (1971) which shows how an employer distaste for a certain group of employees may lower the number of such employees that are hired. The model assumes that all workers are homogeneous and equally materially productive and that the owners of firms are also the employers. However, in Becker's model, the owners of a firm may suffer a cost for each member of their disfavored group that is present in his firm. Because the employer's utility is a function not only of his profits but also of the number of workers from the disfavored group, the employer will be willing to forego some profits to satisfy his discriminatory preferences. This leads us to the following hypothesis:

Hypothesis 1. If employers practice taste-based discrimination, there will be a difference between the hiring rates of male and of female applicants even if they are equally productive.

More specifically, we expect that male applicants will be more likely to be hired by both male and female employers. This is because males are perceived to be more competent than females according to the empirical and experimental work cited in the previous section. We conduct

an Implicit Association Test and a post-experimental questionnaire to understand whether these discriminatory preferences are associated with implicit attitudes or explicitly stated beliefs.

Our second hypothesis deals with the effect of gender quotas among employers in hiring decisions. The effect of having more women or more men in the employer pool, with employers being able to observe who else is an employer, is not obvious. Increased female representation in leadership does not automatically imply that the gender gap in the workplace will narrow. Indeed, female leaders might themselves be biased against female applicants, and male leaders might react to higher female representation by reinforcing discrimination against female applicants. Our experiment is designed to check whether this indeed happens. There is some empirical evidence however that indicates that more female employers lead to more female employees. Cohen and Huffman (2007) use nested data from the 2000 U.S. census to show that the greater representation of women in management narrows the gender wage gap and that the presence of high-status female managers has a signfiicant and large impact on lowering wage inequality. Chambliss and Uggen (2000) find that minority partner representation has a positive effect on minority associate representation. Female decision-makers fill more vacancies with women than do male decision-makers but only among entry-level hires as Gorman (2005) finds in her study using 1990s data from large U.S. law firms. On the other hand, Zinovyeva and Bagues (2010) find that the effect depends on the position at stake. An additional woman in an evaluating committee with seven members increases the number of women promoted to full professor by 14% on average. However, there is no significant relationship between the gender of evaluators and the gender of hires in the case of associate professors.

This leads us to state the following hypothesis:

Hypothesis 2. If gender composition affects hiring decisions, then there will be a difference in the hiring decisions of employers between a male and a female applicant who are equally productive depending on the

gender composition of the pool of employers they belong to.

While the first hypothesis is about the rate of hiring of male and female applicants in the aggregate, the second hypothesis is about the hiring decisions of employers at the individual level. In particular, it deals with how the gender composition of the pool of employers affects the decisions of employers. We do not however specify that male or female applicants are more likely to be hired in pools that have more males or females employers.

We present in the next section our experiment, which was designed to uncover the effect of gender composition on gender bias in hiring while controlling for statistical and taste based discrimination and other sources of discrimination such as other-regarding concerns – which favor members of the group one identifies with – or the wish to compensate for expected discrimination against one's own group by members of the other group. We explain how we did so in the next section.

3. Experimental design

We designed an experiment to test whether discrimination in hiring is taste-based as in Becker's model and whether such discrimination is mitigated by the gender composition of the pool of employer. We mimicked a labor market in the laboratory wherein there are two types of agents—applicants and employers. To lower the impact of statistical discrimination as a possible explanation for preferential hiring, employers in the experiment received direct information about the performance of applicants from both genders. We further discuss in this section details of our design. Instructions given to participants are shown in AppendixA.1.

3.1. Types of participants and tasks

Each session was run with 30 participants, 15 males and 15 females. We assigned 24 participants in the experiment to be employers and 6

to be applicants. Applicants were asked to performed a real-effort task while employers had to perform hiring among applicants. Once this was done, we administered an the Implict Association Test (IAT) and asked participants to fill a post-experimental questionnaire. Table 1 gives a summary of how the experiment proceeded for participants.

Table 1: Overview of Experimental Procedure

Applicant (A)	Employer (B)
Real-Effort Task Round 1 (not remunerated)	Matched with another employer
	Information about applicants Decision who to hire
Real-Effort Task Round 2 (payoff goes to applicant and, if hired this applicant, to the employer)	
	Implicit Association Test Questionnaire

Note: Participants at the beginning of the experiment were assigned either the role of applicant or employer. Applicants first perform a real-effort task. Their performance here feeds into the information that is relayed to employers when they make a series of hiring decision. Applicants perform the same real effort task again which determines their earnings and the earnings of employers. Employers then answer an Implicit Association Test and a Questionnaire.

Real-effort task for Applicants.

Applicants in each session performed two rounds of a real-effort task that consisted in translating letters into numbers within a time limit. A crucial assumption of Becker's model to show the existence of taste-based discrimination is that potential employees in both groups are equally productive in their work. We thus needed a real-effort task that minimizes the differences in the actual performance among applicants across genders and in the beliefs of employers about differences in the competence of applicants at the task because of gender.[1] The

[1]We wanted to avoid distinct gender differences in the attribution of performance like what Deaux and Eimswiller (1974) find, whereby the performance of a male on

decoding task in Kuhn and Villeval (2015) fit these criteria.

Figure A.2 shows a screenshot of the real-effort task. Displayed on applicants' screen was a table with two columns wherein the first column indicated letters and the second column indicated their corresponding numbers. At the right side of the screen was a letter which they have to convert to a number according to the table shown. After they entered a number, they confirmed their answer by pressing OK. A new conversion table was generated only if they correctly converted the letter to the corresponding number.

Based on their performance in a first round of the conversion task, applicants were grouped into three categories: low (first tercile), medium (second tercile), or high (third tercile). This performance grouping, together with their, gender, age, and education was communicated to employers in their hiring task. Whether they were hired or not, applicants performed a second round of the conversion task, which was remunerated 1 ECU (0.1 Euros) for each correctly converted letter.

Hiring task for Employers.

Employers in the experiment made a series of hiring choices between pairs of applicants. We presented 32 hypothetical pairs of applicants along with the 3 real pair of applicants to employers. Employers were informed about each applicants' age, gender, education, and performance in the first round of the real-effort task.[2] We showed information about the applicants' age and education to avoid triggering too much of a demand effect, whereby participants would realize

a task was more likely to be attributed to skill while an equivalent performance by a female on the same task was more likely to be attributed to luck. To check for this, the post-experimental questionnaire asked whether the participants believed males are better at the chosen real-effort task than females.

[2]Employers knew whether an applicant was a low, middle or high performer. To control for considerable differences in employers' interpretation of what low, middle, and high are, we informed them that those assigned to the group low could convert around 40 to 60 letters, those in the group middle could convert around 61 to 80 letters, and those in the group high could convert more than 80 letters. This information was based on a pilot session where another set of participants recruited from the same pool of university students were asked to perform the conversion task.

that the experiment deals with gender discrimination. We also did this to better mimic the actual hiring process wherein employers receive other information than gender about an applicant. Figure A.3 in Appendix A shows a sample of the screen employers encountered. We presented information about applicants jointly in pairs following work by Bohnet et al. (2012) which shows that joint evaluation reduces gender bias compared with separate evaluation. This enabled us to further isolate individual gender preferences as a source of hiring bias instead of other environmental factors.[3]

The task of the employers was to choose which applicant from each pair they would like to perform the same conversion task again for them, *i.e.* the applicant they hired would determine, through its (yet unknown) performance in the second round, the remuneration of the employer (employer and employee were paid the same).

As mentioned, employers encountered 32 hypothetical pairs, the same in all sessions. Mixed in this list were 3 real applicant pairs formed by matching the real participants in the same session. In total, employers encountered and decided for 35 pairs of applicants which were randomly presented across participants and across sessions. Employers did not know which of the pairs they encountered were real or hypothetical. Table 2 presents the characteristics of the hypothetical pairs in terms of gender composition. We balanced age and education in hypothetical pairs across sessions, *e.g.* if in one session applicant pair 31 was a male of age 31 matched with a female of age 35, then in the next session, applicant pair 31 was a male of age 35 matched with a female of age 31.

[3]However, we did not signal gender by using names. This is in contrast to previous hiring experiments wherein names were used to signal membership with a certain group. Rooth (2010), for example, conducted an experiment to test racial discrimination between Swedish- and Arab-sounding names. We avoided using names in this study because names themselves may be associated with certain stereotypes or groups apart from gender.

Table 2: A breakdown of the 32 hypothetical applicant pairs

Applicant pairs	Male is higher	Equal rank	Female is higher	Total
Male-Female	4	12	4	20
Male-Male	4	2		6
Female-Female		2	4	6
Total	8	16	8	32

Note: There were 32 hypothetical pairs of applicants that were shown to participants B. We presented 12 male-female applicant pairs of equal ranks. There were also 8 other male-female applicant pairs, 4 of which had a higher-ranking male and another 4 had a higher-ranking female. There were 12 same-gender applicant pairs, 6 male-male pairs and 6 female-female pairs. Four male-male pairs and four female-female pairs had one higher-ranking member. Two male-male pairs were equally ranked and two female-female pairs were also equally ranked.

Implicit Association Test for Employers.

After the decision task, employers took part in an Implicit Association Test ("IAT") which measured their implicit attitudes on gender (Greenwald et al., 1998). The IAT we implemented is a computer-based sorting task. The IAT consists of 7 blocks. In each, the task of participant is to sort word stimuli according to different categories on the upper portion of the screen. The sorting is executed by pressing either a right or left button. In the gender IAT we implemented, individuals observe words related to male and female, words associated with warmth, and words associated with competence (Fiske et al., 2007). If, for example, one has an implicit view of females being warm, then sorting the words of female and warmth on the same of the screen will be faster than sorting words of female and competence on the same side of the screen. If one has an implicit view of males being competent, then sorting the words of male and competence on the same side of the screen will be faster than sorting words of male and warmth on the same side of the screen. If there is no underlying association, then there should be no difference in the time it takes to sort female and warmth on the same side of the screen and the time it takes to sort female and competence together. A screenshot of the sorting task can be seen in Figure A.4 in Appendix A.

Post-Experimental Questionnaire for Employers.

The last part of the experiment for participant B, we elicited employers' beliefs about gender differences in the experimental task and in other fields. This is to check for overt expressions of prejudice, if there are any. Other items include questions on trust, risk, and demographic characteristics. A copy of the questionnaire can be found in AppendixB.

3.2. Treatments, information, and earnings

We implemented a between-subject design by creating 3 types of employer pairs with varying gender composition: Male with Female, Male with Male and Female with Female. This determined four types of employers: a Male Employer who was paired with a Male Partner ("ME-MP"), a Male Employer who was paired with a Female Partner ("ME-FP"), a Female Employer who was paired with a Male Partner ("FE-MP"), and a Female Employer who was paired with a Female Partner ("FE-FP"). This means that a male participant assigned as an employer can be paired with a male or a female employer. Similarly, a female participant assigned as employer can be paired with a female employer or a male employer. Employers were informed of the gender of their partner in the pool of employers, and the identity (and thus gender) or their partner remained constant for the whole of the experiment (see Figure A.3). Note that employers did not observe the decision of the other employer in their pool, and there was no communication between employers.

All employers underwent the same experimental procedure, the only difference in each treatment was the information they received about the gender identity of the other employer. In total, there were 144 employers in the experiment of which 72 were male and 72 were female. Table 3 summarizes the total number of participants assigned in each pairing condition across the six session of the experiment.

Table 3: Number of Participants, by Gender Composition of Hiring Pair

	Male Partner (MP)	Female Partner (FP)
Male Employer (ME)	36	36
Female Employer (FE)	36	36

Note: There were 144 participants B or employers in our experiment of which 72 were male and 72 were female. We matched participants B into pairs to form a pool of employers. A male participant can be paired another male participant or another female participant. Similarly, a female participant B can be paired with another female participant or a male participant. Each participant B however makes their hiring decisions individually.

As explained, applicants were paid according to their performance in the second round of the conversion task, regardless of whether or not they were hired by an employer, and employers knew this. Employers were paid according to the performance of the applicant they chose in the real pair of applicant that was drawn at random at the end of the experiment. After employers made all their 35 decisions, the computer randomly selected which of the 3 real pairs of applicants who were in the same session would be relevant for their earnings. The same pair was also the one that was relevant for their fellow employer. If both employers in the employer pair chose the same applicant, then the earnings of that applicant in the second round of conversion was also the earning of both employers. If employers hired different applicants in the pair, then one of the employers was selected at random to be the one who determines which applicant was hired, and the performance of that applicant was then the earning of both employers in the pair.

We made clear to all participants that applicants, whether they were selected by an employer or not, would earn the ECUs as per their performance in the second round of conversion task. This was to exclude issues whereby a lower performing applicant would be selected out of pity by an employer. Furthermore, although applicants were aware that their performance in the second round may affect the earnings of an employer, we did not inform them if they were selected by an employer and we did not give them any information about their employer. This was also known to employers.

3.3. *Procedure*

Participants were recruited via ORSEE (Greiner, 2004). Fifteen males and 15 females were invited for each session of the experiment. Six participants (3 males, 3 females) were assigned to be applicants and 24 participants were assigned to be were employers. We conducted six sessions of 30 participants each during the months of July and August of 2014. In sum, the experiment therefore had 180 participants, 144 of whom were employers and 36 of whom were applicants.

We framed the experiment in a neutral way: applicants were referred to as "Participant A" and employers as "Participant B." At the beginning of the experiment, we gave participants a copy of the instructions corresponding to their role (AppendixA.1 and AppendixA.2) and assigned them to a visually isolated computer terminal. All participants had to answer control questions designed to verify their understanding of the experiment. The experiment proceeded once all had answered those questions correctly. The real-effort task for applicants was programmed using Z-Tree (Fischbacher, 2007) while the hiring task and the Implicit Association Test for employers was programmed in E-PRIME (Psychology Software Tools, Inc., 2012).

Participants were 26 years old on average (*SD:* 4.34). Most of them were students from the Friedrich Schiller University or of the University of Applied Sciences in Jena, Germany. Participants received 2.50 euros for showing up on time and earned more depending on their performance and decisions in the experiment. Average payment in the experiment was 11.6 Euros and the experiment lasted one hour to one hour and a half depending on the session.

4. Results

The experiment's main results are as follows. We present the analysis for each result in the subsequent sections.

Result 1. There is no differences in the aggregate hiring rates between male and female applicants. Better-performing applicants get hired

regardless of their gender and the gender of the employer. On pairs of applicants with equal performance but different genders, we observe no difference in the aggregate hiring rates between male and female applicants.

Result 2. The gender composition of the pool of employers influences who gets hired when female employers are paired with male employers. Female employers are more likely to hire female applicants than equally-performing male applicants.

Result 3. Implicit gender bias as measured by the IAT is not associated with hiring decision. Explicit beliefs, however, are associated with the hiring decision.

We begin our analysis of employers and their decisions with the summary statistics in Table 4. An analysis of experimental data from applicants is shown in AppendixC.

4.1. Aggregate hiring rates of male and female applicants

Part 1 of Table 4 shows the average proportion of females hired, by employer, on the hypothetical male-female applicant pairs, which were the same in all treatments. Note from the previous section that each employer encountered 20 hypothetical pairs of this kind. Over all treatments, we observe that employers chose females 51.35% of the time. A chi-squared goodness of fit test shows that this is not significantly different from 50%, ($\chi^2(1, 2880) = 2.11$, $p = 0.15$). We therefore cannot reject the hypothesis that there is no difference in the aggregate hiring rates between male and female applicants. However, as can be seen from the min and max statistics, there were wide variations in hiring rates across individuals: some individuals hired as few as 4 women overall (20% M-F pairs), while others hired as many as 16 women overall (80% M-F pairs).

Table 4: Summary Statistics

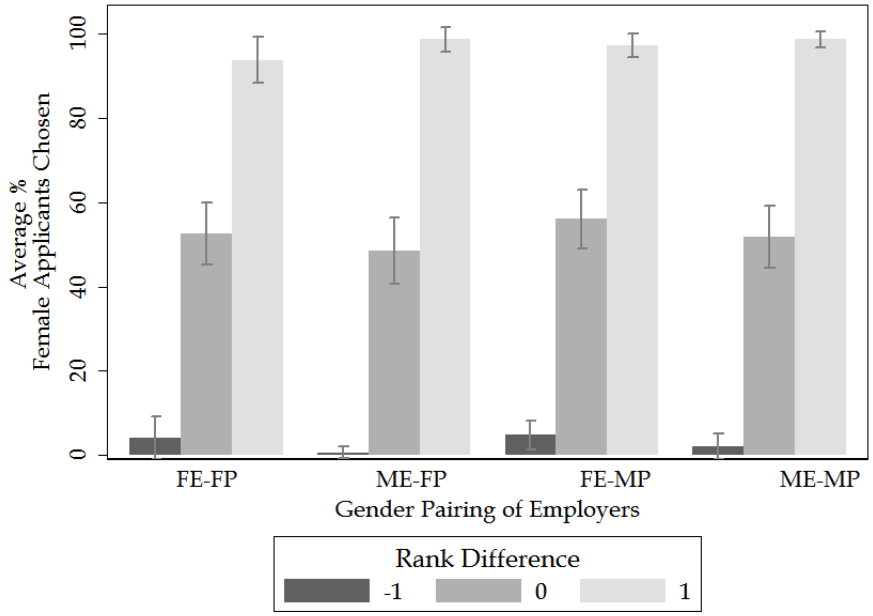
Variable	Mean	SD	Min	Max	N
1. Hiring decisions on M-F Applicant Pairs					
All treatments	51.35	13.49	20	80	144
FE-FP	51.11	13.37	25	80	36
ME-FP	49.08	13.98	25	80	36
FE-MP	54.02	12.97	25	75	36
ME-MP	51.25	13.69	20	70	36
2. Hiring decisions on M-F Applicant Pairs of Equal Performance					
All treatments	52.26	21.73	8.33	100	144
FE-FP	52.55	21.53	8.33	100	36
ME-FP	48.61	23.19	8.33	100	36
FE-MP	56.02	20.66	8.33	91.67	36
ME-MP	51.85	21.74	8.33	83.34	36

Note: The table above shows the summary statistics for the decisions made, IAT scores, and answers to the post-experimental questionnaire. The four treatments are labeled by gender (M for male, F for female) and their role (E as employer, P partner). For example, the hiring decisions of a male employer (ME) matched with a male partner (MP) will be counted under ME-MP. Part 1 is about the percent of female applicant chosen by each employer in the 20 male-female applicant pairs. Meanwhile, Part 2 is the percent of female applicant chosen by each employer in the 12 male-female applicant pairs of equal performance.

4.2. On hiring choice and applicant performance

So far, we have not taken performance into account. We constructed the hypothetical list of applicants to have an equal number of males and females and matched applicants in a manner that differences in performance in male-female applicant pairs cancel out on aggregate. Figure 1 decomposes the hiring decisions per employer on the 20 male-female applicant pairs by rank difference and by gender pairing treatment. Rank was numerically translated so that those with rank "High" had a rank value of 3, those with rank "Middle" had a rank value of 2, those with rank "Low" had a rank of 1. We compute rank difference by subtracting a male applicant's rank from a female applicant's rank. Thus, a rank difference of -1 means that the male applicant had a higher rank, a rank difference of 0 means that both have the

same rank, and a rank difference of 1 means that the female applicant had a higher rank. A glance at the graph shows that rank drives hiring rates. A chi-squared test of independence confirms that there is a statistically significant relationship between performance and hiring rates (Pearson's $\chi^2(2, 2880) = 1.0 \times 10^3$, $p = 0.00$).



Note: The figure above shows the mean percent of female applicants chosen across the four employer pairing conditions. The whiskers indicate the 95% confidence interval of the mean. A rank difference of -1 means that the male applicant had a higher rank, a rank difference of 0 means that both have the same rank, and a rank difference of 1 means that the female applicant had a higher rank.

Figure 1: Mean Percent of Female Applicants Chosen, by Rank Difference and Gender Pairing of Employers

4.3. On hiring choice and gender composition of the pool of employers

The evidence so far shows that aggregate hiring rates do not differ between male and female applicants. We have not yet scrutinized however the effect of our treatment variable, *i.e.* the gender of the other employer. which is where discrimination is the most likely to be discernible. Part 1 of Table 4 shows that male employers chose females roughly half the time (51.25% for ME-MP and 49.08% for ME-FP) while female employers chose females more than 50% of the time (54.02% for

FE-MP and 51.11% for FE-FP). Part 1 of Table 5 shows the results of chi-squared goodness-of-fit tests. We find that the proportion of female applicants that are hired differ from 50% only in the condition when female employers are matched with a male partner. In other conditions, we do not find any statistically significant difference.

Table 5: Chi-squared Goodness-of-Fit Tests on Female Choice in Male-Female Applicant Pairs, by Employer-Partner Pair

	1. All		*2. Equal Performance*	
Employer-Partner Treatment	Pearson's χ^2	p-value	Pearson's χ^2	p-value
FE-FP	0.36	0.55	1.12	0.29
ME-FP	0.27	0.60	0.33	0.56
FE-MP	4.67	0.03	6.26	0.01
ME-MP	0.45	0.50	0.59	0.44

Note: We test the null hypothesis that the percent of female applicant chosen = 50%. We first ran tests on hiring choices on all male-female applicant pairs (N for each treatment is 720) and then on hiring choices on male-female applicants pairs of equal performance (N for each treatment is 432). We report the chi-squared statistic with 1 degree of freedom. Given the results, we reject the null hypothesis that the proportion of female chosen is 50% in the condition where a female employer is paired with a male partner both when deciding for male-female applicants pairs and when deciding for male-female applicant pairs of equal performance. The same results hold if we conduct one-sample binomial tests.

Part 2 of Table 4 deals with the case of equally qualified applicants. The mean proportion of female applicants is not significantly different from 50%, except in the FE-MP condition (56.02%) (Part 2 of Table 5). This result indicates that the gender composition of the pool of employers influences who gets hired when female employers are paired with male employers. Female employers are more likely to hire female applicants, both in the aggregate and in the case of equally-performing male-female applicant pairs.

4.4. On hiring choice and measures of implicit and explicit attitudes

We further build evidence for our first two results in this part by considering in our analysis the effect of explicit and implicit gender attitudes.

Table 6 shows summary statistics of the IAT score and answers to the post-experiment questionnaire. We see that male employers were slower when it comes to associating females with warmth and males with competence than females were (mean response time of 101.89 ms *vs.* a mean response time of 83.70 ms for women). The difference in scores between the two genders is not statistically significant however.

Table 6: Mean IAT Score and Answers to Post-Experiment Questionnaire, by Gender of Employers

	Overalll *N=144*	Male *N=72*	Female *N=72*	*p*-value of two-sided *t*-test
IAT Score (in ms)	92.80	101.89	83.70	0.24
	(91.85)	(105.91)	(74.87)	
Females are better at conversion task	0.29	0.24	0.33	0.20
	(0.45)	(0.43)	(0.47)	
Males are better at conversion task	0.24	0.19	0.29	0.18
	(0.43)	(0.40)	(0.46)	
Deliberately chose females	0.36	0.27	0.44	0.03
	(0.48)	(0.45)	(0.50)	
Deliberately chose males	0.20	0.18	0.22	0.56
	(0.40)	(0.39)	(0.42)	
Male and female performance vary in some fields	0.87	0.85	0.88	0.46
	(0.34)	(0.36)	(0.32)	
Male and female performance differ in medicine	0.20	0.22	0.18	0.53
	(0.40)	(0.42)	(0.39)	
Male and female performance differ in law	0.22	0.25	0.19	0.43
	(0.42)	(0.44)	(0.40)	
Male and female performance differ in science	0.25	0.26	0.24	0.70
	(0.43)	(0.44)	(0.43)	

Note: Standard deviations are reported in parentheses and reported p-values are from two-sided t-tests of independence between male and female values (142 degrees of freedom). IAT Score is measured in milliseconds. All the other variables are binary variables, with a value of 1 for "Yes" and 0 for "No." We find no statistically significant difference in the IAT scores and in the responses between male and female employers in all questions except when asked if they deliberately chose female.

We focus on seven questions in our post-experimental questionnaire that deal directly with employers' gender attitudes. Twenty-nine percent of employers believed that the females are better at the real-effort task in the experiment than males. Meanwhile, 36% of the employers answered that they deliberately chose a female applicant over a male applicant. We also asked the converse of these questions by inquiring if they believed that males are better at the task than females and if they deliberately chose a male applicant over a female applicant. Twenty-four percent of employers answered that they believed females are better at the task than males[4] and 20% of employers said they deliberately chose males over females.

We also asked participants if they believed that there are some fields where males and females differ in performance and questions about specific fields, namely, medicine, law, and science.[5] Eighty-seven percent of employers said they believe that there are some fields where males and female performance differ. 20% answered that there are differences in male and female performance differ in medicine, 22% in law, and 25% in the sciences.

We find no statistically signfiicant difference in the responses between males and females except when we asked if they deliberately chose female applicants. A one-sided t-test reveals that more females than males answered that they deliberately chose a female applicant (p = 0.02). There was no corresponding disproportionate reported bias by male employers for male applicants.

We now consider whether IAT scores and answers in the questionnaire are associated with hiring decisions. Our data consists in a series of decisions by employers over a given set of applicant pairs. We there-

[4]To see whether such beliefs coincide with actual performance of applicants, we find from an analysis of applicants' performance that there is no statistically significant difference in the performance of male and female applicants in the first round of the real-effort task. More details are in AppendixC.

[5]These fields were chosen on the basis of previous research on the existence of hiring discrimination in these fields

fore estimate a random effects logistic regression model as per equation 1 below:

$$y_{ij} = X_{ij}\beta + u_i + \varepsilon_{ij} \tag{1}$$

y_{ij} takes value 1 when employer i chose a female applicant when considering pair of applicants j and X_{ij} is the set of explanatory variables that includes the employer's age, the gender of paired employer, AT score, answers to post-experimental questionnaire, rank difference of the applicant pair, and other control variables. u_i are employer-specific random effect, and ε_{ij} individual decision-specific errors.

Table 7 reports the results of the regressions. We performed 200 bootstrap replications to obtain normal-based 95% confidence intervals for our estimates. This is adequate for normal-approximation confidence intervals (Mooney and Duval, 1993). Regressions confirm what we have found so far in our previous statistical inference. Of all factors in our experiment design, rank difference increases the likelihood of a female applicant being chosen the most.

The IAT score is also not significantly associated with choosing a female applicant. However, employers who believed that females are better at the real-effort task and who answered that they deliberately chose females were more likely to choose females.[6] Interestingly, both these variables are significant, as some participants who believed males were better at the task did not however deliberately choose males, and some who deliberately chose males did not believe that they were better at the task. Employers who answered that they believe that there are differences between male and female performance in some fields and in law and medicine were also less likely to choose female applicants. The 80% of participants who trusted the experimenters to not misuse

[6]We compute the difference between the answer to the question about whether females are better at real-effort task (respectively participant deliberately chose females) and whether males are better at real-effort task (respectively participant deliberately chose males).

their data were also more likely to choose males. There was therefore some demand effect against discrimination for those participants who maybe expected experimenters to be able to relate their decisions to their name.

Table 7: Logistic Regression Estimates of a Model of Choice of Female Applicant in Male-Female Applicant Pairs

	All M-F applicant pairs		Equal ranked M-F applicant pairs	
	β	z	β	z
Characteristics of applicant pairs				
Rank Difference (F - M)	3.79***	(12.94)		
Age Difference (F - M)	0.09***	(3.34)	0.07*	(2.14)
Education Difference (F - M)	-0.60***	(-8.55)	-0.62***	(-7.37)
Employer-Partner Treatments (baseline is ME-MP)				
FE-FP	0.17	(1.11)	0.22	(1.35)
ME-FP	0.11	(0.58)	0.14	(0.69)
FE-MP	0.13	(0.82)	0.12	(0.63)
Individual variables				
IAT Score	0.00	(0.07)	-0.00	(-0.17)
Belief that females are better at task - Belief that males are better at task	0.42**	(2.85)	0.49**	(3.18)
Deliberately chose female - Deliberately chose male	0.62***	(4.93)	0.65***	(5.25)
Male and female performance vary in some fields	-0.52**	(-2.59)	-0.46*	(-2.02)
Male and female performance vary in medicine	-0.37+	(-1.75)	-0.32	(-1.54)
Male and female performance vary in law	0.29+	(1.90)	0.33+	(1.88)
Male and female performance vary in science	0.18	(1.25)	0.21	(1.36)
Belief that experimenters will respect privacy of participants	-0.17*	(-2.25)	-0.22*	(-2.56)
Constant	0.51*	(2.11)	0.50*	(2.02)
Log-Likelihood	-1201.79		-1040.76	
Wald χ^2 (14)	239.86***		180.18***	
N of observations	2880		1728	
N of groups	144		144	

Note: $***$ (<1‰), $**$ (<1%), $*$ (<5%), and $+$ (<10%) mark which variables were statistically significant in each model and indicate their associated levels of significance. The dependent variable is female applicant choice which takes a value of 1 if a female is chosen, 0 if female. z statistics in parentheses are based on bootstrapped standard errors (200 replications)

We find no statistically significant effect of employer-partner treatments on their decision to hire a female applicant in our regressions. However, we could not immediately conclude that the pairing condition does not affect female choice, especially in the light of the results of our previous statistical tests. It is possible that our experimental conditions affect applicant choice indirectly by influencing employers' beliefs

To see if this conjecture holds, we conduct two-sample t-tests to compare whether responses of employers differ from treatment to treatment. We first test the null hypothesis that there is no difference in the responses on the relative competence of females (Belief that females are better at task - Belief that males are better at task) between pairing treatments. Indeed, we could not reject this null hypothesis—responses on these questions do not significantly differ from one treatment to another. This further supports the assumption in our experiment design that the real-effort task is gender neutral in terms of *a priori* beliefs.

We also do similar tests on the responses on the deliberate choice of female applicant over a male applicant (Deliberate choice of female - Deliberate choice of male) between treatments. We find a statistically significant difference in responses only when we compare the ME-FP and FE-MP treatments (t = -2.03, p = 0.04, N=72). It appears that females and males respond differently to being matched with an employer of a different gender. Female employers who are matched with a male partner are not only more likely to choose female applicants as we've seen in previous tests, but are also more likely to say that they deliberately chose a female over a male applicant, at least relative to male employers who are matched with a female partner.

4.5. Discussion

In the two-person pool of employers we formed, we find that female employers who are paired with male employers are more likely to choose a female applicant over an equally competent male. While

our experiment does not fully account for the mechanisms behind this effect, it does help us identify which explanations are more likely. One possible explanation from psychology is ingroup bias or ingroup favoritism (Tajfel and Turner, 1979; Turner and Reynolds, 2010). This is the preferential treatment that is given to others by virtue of having the same group membership. When a female employer is paired with a male employer, one's in-group identity (female) becomes salient in relation to an outgroup which is male. This activates ingroup favoritism, resulting in female employers choosing more female applicants than male applicants.

Ingroup favoritism is stronger when the group is threatened (see Steele and Aronson (1995)). This would explain why only females preferentially chose females when paired with a male employer. As females are less likely to be in decision-making positions in real life, being a female employer paired with a male employer reminds one of one's minority status. This is in contrast to the condition wherein female employers were paired with fellow female employers and both females occupy decision-making positions. As Tajfel and Turner (1979) conceptualize, ingroup bias is an effect that arises from a desire to maintain positive self-esteem or self-worth. When individuals' self-esteem is threatened, as in the case when they are in a discriminated group, individuals turn to group membership to protect their self-worth. The response predicted by the theory is for females paired with male employers to prefer female applicants. Males, who are more likely to be in decision-making positions in real-life, perceive less of a threat from being paired with a female employer and therefore choose females as often as males. That our manipulation on the gender composition of employers affects hiring decisions by females and not by males is also consistent with experimental evidence suggesting that women are more sensitive to social cues (Azmat and Petrongolo, 2014).

Our IAT results show that our employers' implicit attitudes, whether they were male or female, reflect pervading gender stereotypes. Employers more easily associate females with warmth and males with

competence than females with competence and males with warmth. This is consistent with the bulk of past research findings on gender-IAT and IAT in general (Fiske et al., 2007). In our experiment, however, implicit attitude did not translate into a hiring bias while explicit beliefs did. This is perhaps not surprising when we examine hiring as a deliberate process. The message from implicit and explicit literature indicates that explicit attitudes manifest their influence in conscious behavior while implicit attitudes manifest themselves in spontaneous behavior (Dovidio et al. (2002); Jellison et al. (2004); Rydell and McConnell (2006)). In this regard, our results are consistent with this message as we put employers in a situation where their decisions were of direct consequence to themselves and information was available to guide their decisions. Our results also fit with what Sarah Lowes and Weigel (2015) find in their use of IAT to measure ethnic bias in Africa: IAT scores report smaller magnitudes and statistical strength of bias relative to what is found in self-reported views from surveys. It appears that bias tends to be more strongly expressed when self-reported than when measured implicitly. In sum, while we observe preferential hiring for female applicants only when females are paired with males, the effect of gender composition and also implicit attitudes pale in comparison with explicit attitudes when it comes to explaining hiring choice.

5. Conclusion

To say that gender discrimination exists is nothing new. However, disentangling the different types and sources of discrimination (implicit versus explicit, statistical versus taste-based) remains a challenge. To date, previous studies have offered little evidence on what can mitigate hiring discrimination at the level of individual decisions or how perceived discrimination affects choices and behavior. Our work aims to help fill this gap in research and also in the discussion of policies geared towards increasing female representation in decision-making positions.

The relationship between increasing females in decision-making positions and increasing females in entry-level hires is rather complicated given that individuals are influenced by changes in conditions in which they make their decisions. Within the confines of the laboratory, establishing a causal relationship between the increased female representation in the pool of employers and the number of female hires becomes less complex. In this paper, we present a laboratory experiment that depicts a hiring situation wherein markets are competitive and information asymmetries between male and females do not exist. We created two-person employer pools wherein employers received information about an applicant's competence that was directly related to the job that was to be performed. Employers' profits in this experiment were only determined by the applicant hired and they did not incur any other costs in hiring. Applicants not hired could still earn income, hence there was little reason for our participants be concerned about a rejected applicant's welfare. The hiring decision of both employers was also independent, so there was no reason for an employer to favor an applicant to counteract the expected bias of the other employer. We only varied the gender of the other employer in the pool of employers. Under such conditions, the rational decision of an employer faced with the choice between two applicants is to hire the better performer. When making a choice between two equal performers of different genders, there was no reason to prefer one over the other. The only room for discrimination in this experiment was due to individual preferences and prejudices. Such preferences can be linked to implicit gender attitudes and/or explicit beliefs.

Results of our experiment show that female employers when paired with male employers are more likely to choose a female applicant over an equally-competent male applicant. The hiring choices of male employers was not affected by changes in gender composition of the pool of employers. Although employers' implicit attitudes reflected prevailing stereotypes on males being associated with competence and females with warmth, they were not significantly associated with hir-

ing choice. However, their explicit attitudes on gender helped explain hiring choice.

Our experiment provides evidence that observed hiring bias can indeed be partially attributed to individual preferences and influenced by changes in the gender composition of the pool of employers. It should be noted here though that the pool of employers in actual hiring situations rarely receive information about applicants and make hiring decisions the way employers do in our experiment. We eliminated information asymmetries between males and females and between applicants and employers in our experiment, which made discrimination less likely. In spite of our efforts to reduce the possible channels for discriminating behavior, we still found that some participants prefered applicants of one or the other gender. Our result on the role of explicit attitudes points to the role prior beliefs play in discrimination. In spite of being presented information that two applicants were equally productive, those who believe that one gender is better at the task hired someone of that gender more often. It would be interesting to see whether such discriminatory beliefs persist with continued exposure to information on the equal productivity of two groups. To some extent, beliefs about their discriminated status may also have played a role on why females choose more females when paired with a male partner.

Further research

Our experiment laid the groundwork for further experiments that would dissect and separate, as we did, different channels for discriminating behavior to arise. Particularly interesting would be to change some aspects of our design to elicit concerns for the welfare of applicants – by not remunerating applicants that were not hired – and concern about discrimination on the part of the other employer – by taking into account the decision of both employers in choosing who to hire. Our experiment is a good basis for those further variations on our design as we now know how much of discrimination is due to implicit and explicit beliefs, and how little the social context (gender of other

employer), influences hiring decision in our minimal setting.

Further research is also needed in order to explore how broad policies meant to increase female representation in the workforce can address not just market inefficiencies but can also have an impact on individual choices and preferences. Moreover, our study also highlights the importance of understanding how individuals respond to perceived discrimination and to policies meant to address actual discrimination. Our study focused on employers' decisions but further research could also consider applicants behavior. Can changes in the gender composition of management prevent qualified applicants from censoring themselves from applying? Do applicants hired under a quota policy perform any differently than applicants hired under a no-quota policy? Despite the great strides made by women in the last 50 years, it is indeed still vital, and interesting, to further the understanding of how individuals behavior and institutional policies interact to hinder or encourage women's involvement and success in the labor force.

6. References

Ahern, K. R. and A. K. Dittmar (2012). The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation. *The Quarterly Journal of Economics 127*(1), 137–197. 3

Aigner, D. J. and G. G. Cain (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review 30*(2), 175–187. 5

Arrow, K. (1973). The Theory of Discrimination. In O. Ashenfelter and A. Rees (Eds.), *Discrimination in Labor Markets*, pp. 3–33. Princeton, NJ: Princeton University Press. 5

Azmat, G. and B. Petrongolo (2014). Gender and the Labor Market: What Have We Learned from Field and Lab Experiments? Technical report, IZA Discussion Paper No. 8135. 2, 27

Bagues, M. F. and B. Esteve-Volart (2010). Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Random Experiment. *The Review of Economic Studies 77*(4), 1301–1328. 4

Balafoutas, L. and M. Sutter (2012). Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory. *Science 335*, 579–581. 5

Beaman, L., R. Chattopadhay, E. Duflo, R. Pande, and P. Topalova (2009, November). Powerful Women: Does Exposure Reduce Bias? *The Quarterly Journal of Economics 124.4*(4), 1497–1540. 6

Becker, G. S. (1971). *The Economics of Discrimination* (2nd ed.). University of Chicago. 6, 7

Bendick Jr., M., C. W. Jackson, and V. A. Reinoso (1994). Measuring Employment Discrimination Through Controlled Experiments. *The Review of Black Political Economy / 23*(1), 25–48. 5

Bertrand, M., D. Chugh, and S. Mullainathan (2005, May). Implicit Discrimination. *American Economic Review 5*(2), 94–98. 6

Bohnet, I., A. van Geen, and M. H. Bazerman (2012). When Performance Trumps Gender Bias: Joint versus Separate Evaluation. Working Paper RWP12-009, Harvard Kennedy School. 12

Burgess, Z. and P. Tharenou (2002). Women Board Directors: Characteristics of the Few. *Journal of Business Ethics 37*, 39–49. 6

Chambliss, E. and C. Uggen (2000). Men and Women of Elite Law Firms: Reevaluating Kanter's Legacy. *Law and Social Inquiry 25*(1), 41–68. 3, 8

Cohen, P. N. and M. L. Huffman (2007). Working for the Woman? Female Managers and the Gender Wage Gap. *American Sociological Review 72*, 681–704. 8

Darity, W. A. and P. L. Mason (1998). Evidence on Discrimination in Employment: Codes of Color, Codes of Gender. *The Journal of Economic Perspectives 12*(2), 63–90. 5

Deaux, K. and T. Eimswiller (1974). Explanations of Successful Performance on Sex-Linked Tasks: What is Skill for the Male is Luck for the Female. *Journal of Personality and Social Psychology 29*(1), 80–85. 10

Dovidio, J. F., S. L. Gaertner, and K. Kawakami (2002). Implicit and Explicit Prejudice and Interracial Interaction. *Journal of Personality and Social Psychology 82*(1), 62–68. 28

EuropeanCommission (2014). Gender Balance on Corporate Boards: Europe is Cracking the Ceiling. Technical report, European Commission. 2

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics 10*(2), 171–178. 16

Fiske, S. T., A. J. Cuddy, and P. Glick (2007). Universal Dimensions of Social Cognition: Warmth and Competence. *TRENDS in Cognitive Sciences 11*(2), 77–83. 13, 28

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review 104*(4), 1091â€“1119. 2

Gorman, E. A. (2005). Gender Stereotypes, Same-Gender Preferences, and Organizational Variation in the Hiring of Women: Evidence from Law Firms. *American Sociological Review 70*(4), 702–728. 3, 8

Greenwald, A. G., D. McGhee, and J. L. K. Schwartz (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology 74*(6), 1464–1480. 13

Greiner, B. (2004). The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in economics. Working Paper Series in Economics 10, University of Cologne. 16

Huddy, L. and N. Terkildsen (1993). Gender Stereotypes and the Perception of Male and Female Candidates. *American Journal of Political Science 37*(1), 119–147. 6

Inter-Parliamentary Union and UN Women (2015). Women in Politics 2014. 2

Jellison, W. A., A. R. McConnell, and S. Gabriel (2004). Implicit and Explicit Measures of Sexual Orientation Attitudes: Ingroup Preferences and Related Behaviors and Beliefs Among Gay and Straight Men. *Personality and Social Psychology Bulletin 30*(5), 629–642. 28

Krook, M. L. (2006). Reforming Representation: The Diffusion of Candidate Gender Quotas Worldwide. *Politics & Gender, 2*, 303–327. 4

Kuhn, P. and M. C. Villeval (2015). Are Women More Attracted to Cooperation Than Men? *The Economic Journal 125*(582), 115–140. 11

Lim, L. (2009). *Completing the Fertility Transition*, Chapter Female Labour Force Participation, pp. 195–212. United Nations Population Bulletin Special Issue 48/49. 2

Mooney, C. and R. Duval (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference.* Newbury Park, CA: Sage. 23

Moss-Racusin, C. A., J. F. Dovidio, V. L. Brescollc, M. J. Graham, and J. Handelsman (2012, October). Science Faculty's Subtle Gender Biases Favor Male Students. *Proceedings of the National Academy of Sciences of the United States of America 109*(41), 16474–16479. 6

Pande, R. and D. Ford (2012). Gender Quotas and Female Leadership. Technical report, World Bank, Washington DC, USA. 2, 3

Psychology Software Tools, Inc. (2012). E-prime 2.0. 16

Ramirez, F. O., Y. Soysal, and S. Shanahan (1997). The Changing Logic of Political Citizenship: Cross-National Acquisition of Women's Suffrage Rights, 1890 to 1990. *American Sociological Review 62*(5), 735–745. 2

Rooth, D.-O. (2010). Automatic Associations and Discrimination in Hiring: Real World Evidence. *Labour Economics 17*(3), 523–534. 7, 12

Rudman, L. A. and S. E. Kilianski (2000). Implicit and Explicit Attitudes Toward Female Authority. *Personality and Social Psychology Bulletin 26*, 1315–1328. 7

Rydell, R. J. and A. R. McConnell (2006). Understanding Implicit and Explicit Attitude Change: A Systems of Reasoning Analysis. *Journal of Personality and Social Psychology 91*(6), 995–1008. 28

Sarah Lowes, Nathan Nunn, J. R. and J. Weigel (2015). Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test. *The American Economic Review: Papers and Proceedings 105*(5), 340–345. 28

Steele, C. M. and J. Aronson (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology 69*(5), 797–811. 27

Steinpreis, R. E., K. A. Anders, and D. Ritzke (1999). The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study. *Sex Roles 41*(7-8), 509–528. 6

Tajfel, H. and J. C. Turner (1979). An Integrative Theory of Intergroup Conflict. In S. Worchel and W. G. Austin (Eds.), *The Social Psychology of Intergroup Relations*, Chapter An Integrative Theory of Intergroup Conflict, pp. 33–47. Brooks/Cole. 27

Turner, J. C. and K. J. Reynolds (2010). Rediscovering Social Identity: Core Sources. In T. Potmes and N. R. Brandscombe (Eds.), *The Story of Social Identity*, pp. 13–32. Psychology Press. 27

United Nations (2014). Millenium Development Report. Technical report, United Nations. 2

World Bank (2014). World Bank Gender Statistics Database. 2

Zinovyeva, N. and M. Bagues (2010). Does Gender Matter for Academic Promotion? Evidence from a Random Natural Experiment. Technical report, FEDEA. 4, 8

AppendixA. Instructions

Original instructions are in German. We present below a translation of our instructions into English.

AppendixA.1. Instructions for PARTICIPANTS A

Welcome! You are about to participate in an experiment funded by the Max Planck Institute of Economics. Please switch off your mobile and remain quiet. It is strictly forbidden to talk to the other participants. Whenever you have a question, please raise your hand and one of the experimenters will come to your aid.

You will receive 2.5 Euros for showing up on time. Besides this, you can earn more. The show-up fee and any additional amounts of money you may earn will be paid to you in cash at the end of the experiment. Payments are carried out privately, i.e., the others will not see your earnings.

During the experiment we shall speak of ECUs (Experimental Currency Unit) rather than Euros. The conversion rate between them is 1 ECU = 0.1 Euros.

This means that for each ECU you earn you will receive 0.1 Euro.

At the beginning of the experiment, you were given a piece of paper with the password for your computer. Please type this password to activate your screen.

You will then be asked to give your level of education and age. Please be assured that the information you give will not be used to identify you and will be used for experimental purposes only.

In this experiment, there are two types of participants: Participant A and Participant B. You were randomly assigned the role of Participant A.

As Participant A, you will first be asked to convert letters into numbers. Your screen will display a table with two columns. The first column indicates letters and the second column indicates their corresponding numbers. You will be shown a letter and you will have to enter the corresponding number in a box on your screen. You must validate your answer by pressing the 'OK' button.

The conversion table of letters and numbers is modified only when you have correctly entered the correct number. A new conversion table appears when you entered the correct answer.

Please see Figure A.2 for an example of the screen you will encounter in this task.

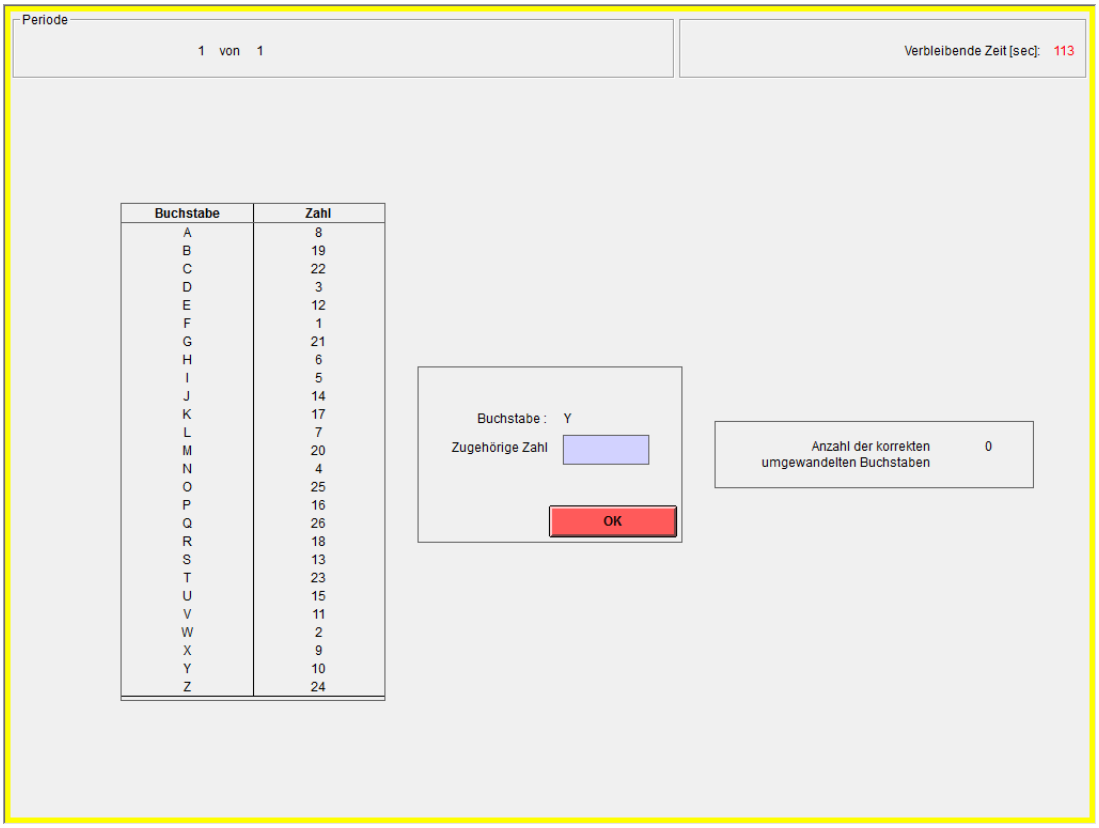


Figure A.2: A screenshot of the conversion task

In this example, you must enter the number 10 and click “OK” for a new conversion table to appear. At the right hand of the screen is the number of correctly converted letters. At the upper right hand of the screen is the remaining time in seconds.

You will be given the opportunity to practice this task during two minutes to familiarize yourself with the task. The number of problems solved during this practice period will not affect your earnings.

You will then go on to perform the task for 4 minutes. Depending on your performance in this round, you will be assigned to one of the following groups: “High”, “Middle,” and “Low.” In the group “High” are one-third of Participants A who converted the most number of letters into numbers. In the group “Low” are one-third of Participants A who converted the least number of letters into numbers. All remaining participants A are then assigned to the group “Middle”

The points you earned in this first round of conversion is not relevant for your earnings.

After completing this task, participants who were assigned the role of Participant B for this experimental session will be shown your group based on your performance in the first round of conversion, your level of education, and your age.

Participants B will decide whether they want your performance in a second round of this tasks to determine their payoff or not.. If you are selected, then whatever you earn in the second round is also what participants B will earn. So if you earn XXX ECU in the next round then participants B will also earn XXX ECU. You will not be informed of whether you were selected or not.

Whether you are selected by a Participant B or not, you will be able to perform the conversion task once again for 4 minutes and you will be paid for each correct conversion. For each letter you correctly convert to a number, you will get 1 ECU. The more letters you convert correctly into a number, the more ECUs you will get. Your payment does not decrease if you give an incorrect answer to a problem, but you are not paid for incorrect answers.

Please click OK on the screen when you are done reading the instructions. You will then be asked to answer some questions to check your understanding of the instructions. The experiment will go on only once all participants answered all questions correctly.

AppendixA.2. Instructions for PARTICIPANTS B

Welcome! You are about to participate in an experiment funded by the Max Planck Institute of Economics. Please switch off your mobile and remain quiet. It is strictly forbidden to talk to the other participants. Whenever you have a question, please raise your hand and one of the experimenters will come to your aid.

You will receive 2.5 Euros for showing up on time. Besides this, you can earn more. The show-up fee and any additional amounts of money you may earn will be paid to you in cash at the end of the experiment.

Payments are carried out privately, i.e., the others will not see your earnings.

During the experiment we shall speak of ECUs (Experimental Currency Unit) rather than Euros. The conversion rate between them is 1 ECU = 0.1 Euros.

This means that for each ECU you earn you will receive 0.1 Euro.

The experiment is composed of two parts. Only the first part is relevant for your payment. The instructions for the first part follow shortly. The instructions for the second part will be shown on your screen when you are finished with the first part.

AppendixA.2.1. Part 1

In this experiment, there are two types of players: Participant A and Participant B. You were randomly assigned the role of Participant B.

At the beginning of the experiment, you were paired with another Participant B to form a pair. You and your fellow Particpant B in the pair will be shown the same pair of Participants A. Participants A have performed a task of converting letters into numbers based on a table of correspondence between letters and numbers. They had 2 minutes to practice this task. They then performed the same task for 4 minutes. Participants A have been assigned according to their performance in this 4-minute round in one of the following groups: High, Middle, and Low.

In the group "High" are one-third of Participants A who converted the most number of letters into numbers. In the group "Low" are one-third of Participants A who converted the least number of letters into numbers. All remaining participants A are then assigned to the group "Middle" Based on the performance of participants who already performed this task in a previous session, those assigned to the group low can convert around 40 to 60 letters, those in the group middle can convert around 61 to 80 letters, and those in the group high an convert more than 80 letters.

You and the other Participant B in your group will decide which Participant A will perform the conversion task again and determine your earnings.

The performance of the selected Participant A determines your earnings and the earnings of the other Participant B in your group and also the earnings of Participant A. The emans, for each letter that Participant A correctly converts in the second round of the task, you and Partcipant A receive 1 ECU.

Your task is to decide which of Participant A in the pair will perform the conversion task for another round of 4 minutes for you. The other Participant B encounters the same pairs as you but you will make the decisions individually. This means you will not know of the decision of the other Participant B. Participants A are also not informed whether they are selected or not.

Below is a sample of the decision screen you will encounter.

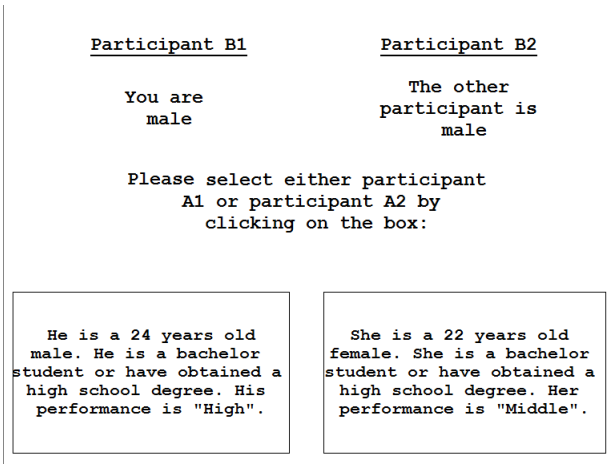


Figure A.3: A sample of the decision screen for Participant B

In the screenshot above, you are Person B1 and the other participant in your group is Person B2. Your respective genders are also shown. Below this, you can see the two Participants A, Person A1 and Person A2. Apart from their performance in the first round of the conversion task, you will also be informed of each Participant A's age and education level. Please click on which Participant A you would

like to determine your earnings in a second round of conversion task. When you have made your choice, another pair of Participants A will be shown. You will encounter each pair and each Participant A only once.

Overall, you will be shown 35 different pairs of Participants A to select from during the course of the experiment. Some of these participants are hypothetical, others are real.

After having made those 35 decisions, the computer will randomly select which of the pairs of real Participants A that you were shown will be relevant for your earnings. The same pair will also be relevant for your fellow Player B. Because you do not know which pairs will be relevant for your pay-off, you should decide as if each pair was the one that will be chosen to be payoff relevant.

Then the computer chooses randomly whose decision in the relevant pair will be taken into account – yours or that of your fellow Player B. That is, with probability one-half the Participant A whom you chose will determine your earnings. Otherwise, it is the other Participant A who will determine your earnings. This means that your choice counts in only half of the cases. You will be informed at the end which pair of Participant A and which Participant A in the pair was chosen. Because your decision has the same chance to count as that of the other, please decide as if you are the one who will determine which Participant A is selected.

All Participants A will perform the conversion task again for 4 minutes. The Participant A who was selected according to the procedure above will NOT be told about your choice. Both you, your fellow player B and this player A will receive a payoff equal to this player's performance in this second round of the conversion task.

The Participant A who was not selected will also perform the conversion task again and receive a payoff equal to their performance, 1 ECU for each correctly converted letter. That performance does not determine your payoff however, and you will not be told how other Particpants A performed.

You will then be asked to answer some questions to check your understanding of the instructions. After answering the questions, your decision task will begin. To proceed, kindly click on the file "ParticipantB.ebs" that will appear on your screen. The folder where this file can be found is already open. Please wait a moment in case the file is not yet shown on your screen.

When the file appears on your screen, please double-click to open. Upon entering the laboratory, you were given a small sheet of paper with your participant number, cabin number, and password. Kindly enter this information on the screen when it is requested.

You will then be requested to enter your education level and age. Please be assured that the information you give will not be used to identify you and will be used for experimental purposes only.

AppendixA.2.2. Part 2: Implicit Association Test

We presented instructions for the IAT on the screen. Below is a sample of the screen participants encountered. In the middle of the screen is a word that a participant sorts to one of the categories at the right and left side of the screen by pressing a right and left button respectively. If, for example, one has an implicit view of women being warm, then sorting the word women with female and warmth on the same of the screen will be faster than sorting it with words of female and competence on the same side of the screen. If one has an implicit view of men being competent, then sorting the word men with the words male and competence on the same side of the screen will be faster than sorting it with the words male and warmth on the same side of the screen.

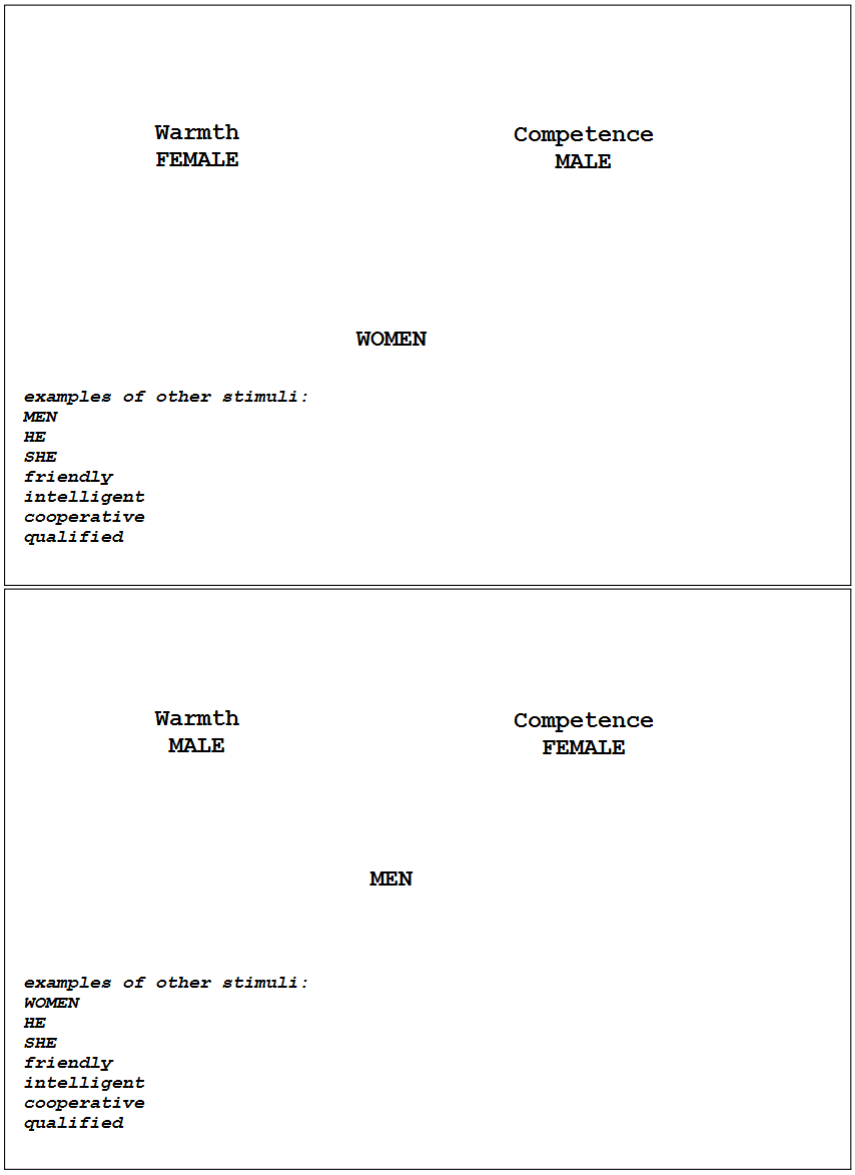


Figure A.4: Screenshots of the Implicit Association Test

AppendixB. Post-Experimental Questionnaire

English translation of the post-experimental questionnaires:

1. Which factors did you take into consideration when you were making your decision? (Gender / Age / Education / Combined Factors).
2. Did you believe that females perform better than males in Player A task? (Yes/ No).

3. Did you believe that males perform better than females in Player A task? (Yes/ No).
4. Did you deliberately choose female participant over male participant? (Yes/ No).
5. Did you deliberately choose male participant over female participant? (Yes/ No.)
6. Did you believe that younger people perform better than older people in Player A task? (Yes/ No).
7. Did you believe that higher educated people perform better than lower educated people in Player A task? (Yes/ No).
8. Would you have decided differently, if the gender of your fellow Player B had been different than what it was? (Yes/ No).
9. What was the gender of your fellow Player B? (Male/ Female).
10. Did you expect that your fellow Player B would select Player A with the same gender (as the fellow player B)? (Yes/ No). If yes, why?
11. Did you expect that your fellow Player B would select Player A with the opposite gender (as the fellow player B)? (Yes/ No). If yes, why?
12. Did you expect that your fellow Player B would take the gender of Player A into consideration? (Yes/ No). If yes, why?
13. Do you believe that there are areas in which the performance of females and males differ? (Yes/ No).
14. In the field of medicine? (Yes/ No). If yes, in which way do they differ?
15. In the field of law? (Yes/ No). If yes, in which way do they differ?
16. In the field of science? (Yes/ No). If yes, in which way do they differ?
17. In other fields? (Yes/ No). If yes, in which field and what are the differences?
18. What do you think was the purpose of the experiment?
19. How difficult did you find it to come up with a decision during the experiment? (7-point scale from “very easy” to “very difficult”).

20. How understandable did you find the instructions of the experiments? (7-point scale from “very understandable” to “very difficult to understand”).
21. Did you know any of the other participants of this experiment? (Yes/ No). If yes, how many people did you know?
22. Did you have any problems during the experiment? (Yes/ No). If yes, what are they?
23. Did you find the payment of this experiment appropriate? (Yes/ No).
24. Do you believe, that the experimenter will not misuse any of your personal data from this experiment? (Yes/ No).
25. What is your nationality? (German or _____).
26. Please indicate your current activities (e.g. studying, working)? _____ (maximum of 2 entries).
27. Are you currently a student? (Yes/ No). If yes, in what field are you currently studying?
28. Where did you live (most of your life)? (Big city with more than 1 million peope/ Big city with more than 100.000 people/ City with more than 10.000 people/ A village/ Others: ____).
29. What are the main sources of your finances (e.g. family, scholarship, salary, government help/loan)?
30. How risk-taking are you in general? (Please give a number between 0 and 10. Zero for avoiding as much risk as possible and 10 for being very risk-loving).
31. Do you believe that two-people with the same qualifications should be paid equally, even though one person is more productive than the other? (Yes/ No/ Not Sure).
32. What do you think of the following statement? In general, people can be trusted. (7-point scale from “totally agree” to “do not agree at all”).
33. What do you think of the following statement? Nowadays, we cannot trust on people so easily. (7-point scale from “totally agree” to “do not agree at all”).

34. What do you think of the following statement? When you are dealing with stranger, it is better to be careful before you put your trust into that person. (7-point scale from "totally agree" to "do not agree at all").
35. Do you believe that most people will take advantage of you when there are opportunities? (7-point scale from "I very much believe" to "I do not believe at all").
36. Do you believe that most people treat others fairly? (7-point scale from "I very much believe" to "I do not believe at all").
37. Would you say that people most of the time strive to be helpful to others? (Yes/ No).
38. Would you say that people most of the time strive only to fulfill their own interest? (Yes/ No).

AppendixC. Analysis of Applicant Performance

In each session, 6 participants (3 males, 3 females) were assigned as participant A or applicants. We ran 6 sessions giving us a total of 36 applicants. The average age of applicants in the experiment was 25.91 years (*SD 7.38*). Applicants who were studying for their Bachelor degree comprised 30.58% of the total, 52.78% of the applicants were studying for their Master degree, 5.56% of the applicants were doing their PhD degree, and 11.11% of the applicants listed "other" as their education level.

We first check whether there are signficant differences in the mean and variance of performance between male and female applicants and between the two rounds. C.8 shows how applicants performed by gender and round and results of t-tests of differences in means. We then compare whether there are differences in the first round and second round of performance. We expect that performance in the first round will be lower than the second round because the first round is not paid and because by the second round, participants are now more familiar with the task. We therefore conduct a one-sided t-test and find that

males and females perform better in the second round.[7] The increase in performance is slightly higher for males than females.

We now go to the more relevant comparison for our study. We assume that the task is gender-neutral and have no strong *a priori* reason to believe why males or females should perform better than the other. We thus use a two-sided t-test difference of means. If we compare performance between males and females in the first round, we find no statistically signficant difference in performance. Once we compare performance between males and females in the second round which is paid, we find a statistically significant difference.

Table C.8: Mean Performance of participants A, by Gender and Round (Standard Deviations in Parentheses)

	Male	Female	p-value from 2-sided t-test)
First Round	99.89 (13.46)	92.78 (17.53)	0.181
Second Round	107 (15.07)	96.39 (15.65)	0.0460
p-value from one-sided t-test	0.0230	0.073	

Note: The p-values reported are from t-tests checking for differences in mean performance. The rows compare mean performance between gender and difference in mean performance between males and females while the columnc compare mean performance across rounds. We find that differences in mean performance between males and females is not statistically significant in the first round but is statistically significant in the second round according to 2-sided t-tests, i.e., males perform better than females in the second round. Both male and female participants show higher mean performances in the second round according to results of one-sided t-tests.

[7] A two-sided t-test looking at the difference of first and second round performance for males yields a p-value of 0.046; for females yields a p-value of 0.1461.