Barron, Kai; Ditlmann, Ruth; Gehrig, Stefan; Schweighofer-Kodritsch, Sebastian

Working Paper

Explicit and implicit belief-based gender discrimination: A hiring experiment

WZB Discussion Paper, No. SP II 2020-306

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Barron, Kai; Ditlmann, Ruth; Gehrig, Stefan; Schweighofer-Kodritsch, Sebastian (2020) : Explicit and implicit belief-based gender discrimination: A hiring experiment, WZB Discussion Paper, No. SP II 2020-306, Wissenschaftszentrum Berlin für Sozialforschung (WZB), Berlin

This Version is available at:
https://hdl.handle.net/10419/223247

WZB

Wissenschaftszentrum Berlin
für Sozialforschung

Kai Barron
Ruth Ditlmann
Stefan Gehrig
Sebastian Schweighofer-Kodritsch

Explicit and implicit belief-based gender discrimination: A hiring experiment

Discussion Paper

SP II 2020–306

August 2020

Research Area
Markets and Choice

Research Unit
Economics of Change

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Affiliation of the authors:

Kai Barron, WZB (kai.barron@wzb.eu)

Ruth Ditlmann, WZB (ruth.ditlmann@wzb.eu)

Stefan Gehrig, WZB, (stefan.gehrig@wzb.eu)

Sebastian Schweighofer-Kodritsch, Humboldt-Universität zu Berlin
(sebastian.kodritsch@hu-berlin.de)

Abstract

Explicit and implicit belief-based gender discrimination: A hiring experiment[*]

Understanding discrimination is key for designing policy interventions that promote equality in society. Economists have studied the topic intensively, typically taxonomizing discrimination as either taste-based or (accurate) statistical discrimination. To enrich this taxonomy, we design a hiring experiment that rules out both of these sources of discrimination along the gender dimension. Yet, we still detect substantial discrimination against women. We provide evidence of two forms of discrimination, explicit and implicit belief-based discrimination. Both rely on statistically inaccurate beliefs but differ in how clearly they reveal the decision-maker's gender bias. Our analysis highlights the central role played by contextual features of the choice environment in determining whether and how discrimination will manifest. We conclude by discussing how policy makers may design effective regulation to address specific forms of discrimination.

Keywords: Discrimination; Hiring Decisions; Gender; Beliefs; Experiment

JEL classification: D90, J71, D83

[*]The authors would like to thank Thomas Graeber, Jon de Quidt, and the audience at the WZB Brown Bag Seminar for helpful comments. We thank the WZB for generously funding this project through means of its interdisciplinary "seed money" programme. Barron and Schweighofer-Kodritsch gratefully acknowledge financial support by the *Deutsche Forschungsgemeinschaft* through CRC TRR 190.

1 Introduction

Discrimination in the labor market is a critically important policy issue around the world.[1] When one individual receives preferential treatment over another on the basis of gender or ethnicity, this violates basic meritocratic principles. It is also inefficient if it results in a less productive workforce due, for instance, to (i) lower returns to educational investments for some groups, or (ii) sub-optimal matching between skills and tasks. Moreover, such discrimination predominantly harms socio-economically weaker groups and thereby reinforces inequality. In relation to gender, which is the focus of this paper, a substantial body of work has provided evidence that discrimination plays an important role in generating the gender gap observed in wages and career progression.[2] However, in addition to being able to detect discrimination it is crucial for the design of effective policies to be able to understand the underlying causes of discrimination.

Traditionally, the economics literature has distinguished between discrimination based on taste (Becker, 1957) and discrimination resulting from rational beliefs (Phelps, 1972; Arrow, 1973). In the former, an employer is willing to pay a price to avoid transactions or being "associated with some persons instead of others" (Becker, 1957, p. 14). In the latter, information about true group differences in productivity leads to discrimination. However, recent work suggests that this taxonomy may be too narrow in important ways. First, it emphasizes the prevalence of discrimination emanating from "irrational" (i.e., inaccurate) beliefs due, for example, to widely held stereotypes (see, e.g., Judd and Park, 1993; Hilton and Von Hippel, 1996; Bordalo, Coffman, Gennaioli, and Shleifer, 2016; Bohren et al., 2020). Second, building on insights from social psychology (e.g., Snyder, Kleck, Strenta, and Mentzer, 1979; Banaji and Greenwald, 1995; Hodson, Dovidio, and Gaertner, 2010), it highlights a tension between appearing non-discriminatory to oneself or others and actually holding discriminatory stereotypes/preferences (see Bertrand, Chugh, and Mullainathan, 2005; Bohnet et al., 2015; Cunningham and de Quidt, 2016; Bertrand and Duflo, 2017; Danilov and Saccardo, 2017; Carlana, 2019). In cognitively difficult or ambiguous choice settings, an individual facing a tension of this nature is prone to discriminate *implicitly*, con-

[1] For a review of the economics literature on discrimination, see: Riach and Rich (2002), Charles and Guryan (2011), Lane (2016), Bertrand and Duflo (2017) and Blau and Kahn (2017).

[2] Evidence has been documented in a diverse range of contexts including *bargaining* (Ayres and Siegelman, 1995; Bowles, Babcock, and Lai, 2007; Small, Gelfand, Babcock, and Gettman, 2007), *hiring* (Jowell and Prescott-Clarke, 1970; Newman, 1978; McIntyre, Moberg, and Posner, 1980; Yinger, 1986; Riach and Rich, 1987; Glick, Zion, and Nelson, 1988; Neumark, Bank, and Van Nort, 1996; Biernat and Kobrynowicz, 1997; Goldin and Rouse, 2000; Bertrand and Mullainathan, 2004; Reuben, Sapienza, and Zingales, 2014; Bohnet, Van Geen, and Bazerman, 2015; Milkman, Akinola, and Chugh, 2015; Kübler, Schmid, and Stüber, 2018; Bohren, Haggag, Imas, and Pope, 2020; Coffman, Exley, and Niederle, 2020), *referrals, promotions, and recognition for (group-)work* (Isaksson, 2018; Coffman, Flikkema, and Shurchkov, 2019; Sarsons, 2019; Hengel, 2020; Card, DellaVigna, Funk, and Iriberri, 2020; Sarsons, Gërxhani, Reuben, and Schram, 2020).

tradicting the beliefs/preferences she expresses *explicitly*. Such implicit discrimination falls outside the scope of the (subjective) expected utility framework on which all of the aforementioned explanations are based. It is a particularly problematic form of discrimination because: (i) by it's nature, it is even more difficult to identify, and therefore regulate, than explicit forms of discrimination are, and (ii) it is likely to materialize in precisely the contexts where explicit discrimination has already been acknowledged to be unethical and is therefore highly stigmatized.

In this paper, we study both explicit and implicit gender discrimination using a stylized labor market hiring experiment. The design allows us to completely rule out taste-based discrimination and focus on isolating different forms of belief-based discrimination. In this, we join the contemporary literature considering the central role of (possibly inaccurate) beliefs in generating different behavior towards men and women (see, e.g., Bordalo et al., 2016; Bohren, Imas, and Rosenberg, 2019; Coffman et al., 2019; Bordalo, Coffman, Gennaioli, and Shleifer, 2019; Bohren et al., 2020).[3] We therefore also ask whether any implicit or explicit discrimination we observe is consistent with statistically accurate beliefs. Further, our data allows us to present evidence on whether image concerns (i.e., a fear of appearing sexist) play a role in influencing our employers' decisions in the form of excuse-driven behavior (see, e.g., Exley, 2016; Danilov and Saccardo, 2017; Coffman et al., 2020).

We designed a tightly controlled experiment that simulates the main features of a hiring scenario, but allows us to study these different manifestations of discrimination. This requires making careful adjustments to the information environment in which participants make their hiring decisions. In the experiment, a first group of 80 participants serve as job candidates, and a second group of 240 participants take on the role of employers. These employers make a series of anonymous hiring decisions between pairs of candidates. For each candidate, the employer receives a "mini-CV" that provides information about gender and two possible qualifications. A qualification takes the form of a "certificate" that is awarded to candidates who score in the top 30% in a particular qualification task. There are two qualification tasks – a general knowledge task and a word search task. These qualification tasks are distinct from the job task (a logic task). Employers are incentivized to hire the better job candidate – they receive a fixed payment if they hire the candidate that performs better in the job task. Structuring the incentives in this way achieves multiple objectives. First, the fixed payment (as opposed to paying the employer in proportion to the candidate's output) ensures that the employer is incentivized to choose the candidate they believe is most likely to be better; thus, we remove the role of risk preferences as well as avoiding an exces-

[3]Aside from this recent wave of papers, Bohren et al. (2020) note that very little empirical work in economics between 1990 and 2018 considered the role of inaccurate beliefs in discrimination. Their review of the literature indicates that only 4.8% of the 105 papers they identified in top economics journals tested for inaccurate beliefs. Collectively, the recent literature suggests this is an important omission.

sive influence of high- (or low-) performing outliers. Second, this incentive structure allows us to focus on belief-based forms of discrimination by making the hiring decision inconsequential for the candidates; they never learn about the decision nor is their pay affected by it. This rules out by design any gender discrimination based on concern for others' payoffs (material or psychological) or based on transactions with certain groups, differentiating our work from other experimental work on gender discrimination such as that of Bohren et al. (2020) or Reuben et al. (2014), where the evaluator's/employer's choices directly affect the candidate's fate.[4]

To identify discrimination, we compare decisions made by employers between real candidates with "qualification profiles" a and b, say, where in one scenario a belongs to a female and b to a male candidate, and in another scenario this is reversed. Any significant difference in how often qualification profile a gets chosen between the two scenarios can be cleanly attributed to the gender of the candidates and constitutes gender discrimination.[5]

We find the following patterns of belief-based gender discrimination: First, whenever both candidates are *equally qualified* (i.e., both have identical qualification profiles), there is significant discrimination against women. Since such decisions are only about gender, and cannot be attributed to any other candidate characteristics, we term this "explicit" discrimination. Here, it is "statistically inaccurate" (we borrow this terminology from Bohren et al., 2020), since female candidates were not objectively out-performed by male candidates; in this sense, our experiment rules out (accurate) statistical discrimination. Second, whenever one profile is *more qualified* than the other (i.e., one has a certificate, the other has none), gender plays no role in hiring, i.e., there is no discrimination.

However, third, whenever both profiles are *differently qualified* (i.e., each candidate has a different certificate), there is again significant discrimination against women. The magnitude of the gender gap here is almost as large as for the decisions where employers have no information upon which to differentiate the candidates other than their gender. This could simply mean that gender is perceived as a more reliable signal of productivity in the job task than the certificates, of course, but intuitively it does not square well with the complete absence of discrimination when one has a certificate and the other does not. Indeed, combining the between- and within-subject choice patterns in our data, we identify a significant fraction of employers that always hire a female candidate over an equally qualified

[4]Though these features of the hiring choice may seem somewhat artificial with respect to actual labor market discrimination, it is quite plausible that there are often hiring committees including members that (i) have an incentive (potentially intrinsic) to hire the best candidate for the organization, (ii) won't ever interact with whoever gets hired or not after the completion of the hiring process, and (iii) do not care about any of the candidates' outcomes directly. If gender enters such a committee member's evaluation, we would still call this gender discrimination.

[5]Strictly speaking, we thereby measure gender discrimination in *relative* terms as more discrimination against one gender in comparison to the other.

male candidate (where discrimination is obvious), yet would always hire a male candidate over a female candidate that is differently qualified, *regardless of what these qualifications are* (where discrimination is more opaque). In other words, such employers are found to discriminate only implicitly. This corresponds to an intransitive cycle and could therefore not be rationalized by any well-defined preference/beliefs over candidates (see Cunningham and de Quidt, 2016). Notably, we find no such implicit belief-based discrimination with respect to any other attributes of the CV.

The pattern of behavior that we observe is consistent with a stereotype about male superiority in logic tasks leading employers to form gender-biased beliefs (see also Bordalo et al., 2019). In fact, by eliciting the beliefs of the job candidates themselves, we find that they believe that men perform better in the logic task (i.e., the job task). The implicit discrimination that we see could therefore be due to a combination of: (i) an aversion to displaying overtly discriminatory behaviour, and (ii) underlying (mistaken) stereotypes. This aversion amongst some subjects to overtly discriminate against women may emanate from the fact that gender discrimination against women is stigmatized in certain segments of the population (our subject pool is comprised of young, highly educated Westerners). Despite any beliefs they may hold about who is the better candidate, employers may wish to signal that they do not discriminate. While we cannot unequivocally demonstrate that such self or social image concerns drive implicit discrimination in our experiment, additional features of our experimental design allow us to study whether employer behavior is consistent with this explanation. We indeed obtain suggestive evidence of "covering up" the intent to preferably hire a man; in particular, the beliefs data we elicit about the relative importance of the two potential qualifications for job performance suggests that employers rationalize gender-biased choices as being qualification-driven by "redefining merit" ex post (see, e.g., Uhlmann and Cohen, 2005). This involves adjusting one's beliefs about the relative importance of each of the qualifications for predicting performance in the job task to justify one's gender-based decisions.

Taken together, our results show evidence of both explicit and implicit gender discrimination, based on statistically inaccurate gender stereotypes. There are several important implications of the evidence reported in this paper. First, our results demonstrate that discrimination can take very different forms beyond the traditional distinction of taste-based and (accurate) statistical discrimination. This is important because if we wish to design policy interventions that effectively combat discrimination, we need to be able to recognize all its manifestations. Further, to choose the correct policy instrument to address discrimination in a particular context, it is imperative to understand the root cause of the problem.[6]

[6]Bohren et al. (2020) also provide an insightful discussion of the importance of correctly identifying the source of discrimination in order to design an effective policy intervention to address it.

Otherwise, the treatment may be ineffective or lead to undesired consequences. For example, policy makers that wish to fight discrimination may be tempted to impose rules that address explicit discrimination, such as: "if choosing between an equally qualified man and woman, choose the woman."[7] While this may be effective in some contexts, in most real-world hiring decisions the candidates differ on many dimensions, so that the evaluation of candidates' overall suitability for a position depends on the rather malleable and subjective relevance assigned to their different attributes. In such contexts with highly heterogeneous candidates, an affirmative action policy rule of this type might be ineffective for changing hiring behavior, while creating the illusion that discrimination is being addressed. Further, it may also lead to individuals going to greater lengths to mask or obscure their discriminatory decisions (even possibly from themselves).

Second, the manifestation of discrimination (both its occurrence and its form) depends crucially on the choice setting. In the same pool of subjects, we document evidence of discrimination when candidates are *differently qualified*, but not when one candidate is *more qualified*.[8] This suggests that discrimination is more likely to occur in settings where candidates are heterogeneous on multiple job-relevant attributes (horizontal heterogeneity), and less likely to occur when candidates are heterogeneous on a single dimension (vertical heterogeneity). This has meaningful implications for the design of remedies that involve altering the architecture of the choice environment. In particular, situations with horizontal heterogeneity can be translated into situations with vertical heterogeneity by means of carefully designed procedures. For example, one potential solution is to require ex ante criteria that specify how to evaluate candidates on different dimensions, and how to aggregate these evaluations into a single score, as is sometimes already the case in university acceptances. This would avoid the ex post re-weighting of the importance of different attributes (e.g., as discussed in Hodson, Dovidio, and Gaertner, 2002 and Uhlmann and Cohen, 2005).

Third, our results speak to the long history of theories of human behavior that posit a tension between hidden and expressed motives – ranging from Freud to the modern widespread usage of the implicit association test (IAT) in social psychology (Greenwald, McGhee, and Schwartz, 1998). One of the key tensions studied in this social psychology literature is the underlying conflict between explicit egalitarian beliefs and implicit racial biases (Hodson et al., 2010). More recently, the IAT has been used as an effective tool for studying the influence of implicit stereotypes in the economics literature. For example, Carlana (2019) shows that the gender stereotypes of teachers can have a substantial impact on the out-

[7] Cunningham and de Quidt (2016) refer to these as *ceteris paribus* rules. We will use a more general terminology and refer to them as "affirmative action rules".

[8] In addition, we also document evidence of discrimination when candidates are *equally qualified*. This indicates that discrimination is also an issue when candidates are highly homogeneous (in terms of their suitability for the job).

comes of their students (increasing the gender gap in math performance, and inducing girls to self-select into less ambitious high school tracks).[9] The IAT aims to assess the strength of associations between concepts (e.g., "female" / "male" and "logic") by measuring response times when subjects classify concepts together in a computerized task. Here, we demonstrate a complementary approach to eliciting implicit preferences, following the proposed methodology of Cunningham and de Quidt (2016).[10] Unlike the IAT which relies on response times as a proxy measurement, we measure both explicit and implicit preferences from actual choice data. Since the efficacy of the IAT in predicting real-world discrimination is still a contentious topic (see, e.g., Oswald et al., 2015 and Kurdi et al., 2019), a behavioral measure as presented here provides an instructive complement to the IAT. Since implicit preferences are by their nature difficult to detect, it is useful to have different measurement tools, each of which may be more appropriate for a particular subset of research contexts.[11]

The paper proceeds as follows. Section 2 describes the experimental design. Section 3 presents the results, while Section 4 discusses the policy implications and concludes. The Appendix provides additional details.

2 Methodology

2.1 Experimental Design

The experiment consisted of two parts, each conducted with a separate group of subjects. In the first part, the Job Candidate Assessment experiment, we collected information from 80 subjects regarding their performance in several tasks – a general knowledge quiz (qualification task 1, which we refer to as the *knowledge task*), a word search puzzle (qualification task 2, which we refer to as the *word task*), and a matrix logic exercise (the job task, which we refer to as the *logic task*). Each of the subjects also self-reported the gender that they identified with. These individuals served as our *job candidates* during the main part of the

[9]The IAT has also been used to study implicit racial or ethnic bias by, e.g., Rooth (2010), Glover, Pallais, and Pariente (2017), Corno, La Ferrara, and Burns (2019) and Alesina, Carlana, La Ferrara, and Pinotti (2018). The results in Alesina et al. (2018) highlight the immense importance of both: (i) knowing how to detect different forms of discrimination, and (ii) tailoring the policy response to the specific form of discrimination. In their study, teachers who are simply made aware of their implicit biases reduce their discriminatory grading behavior.

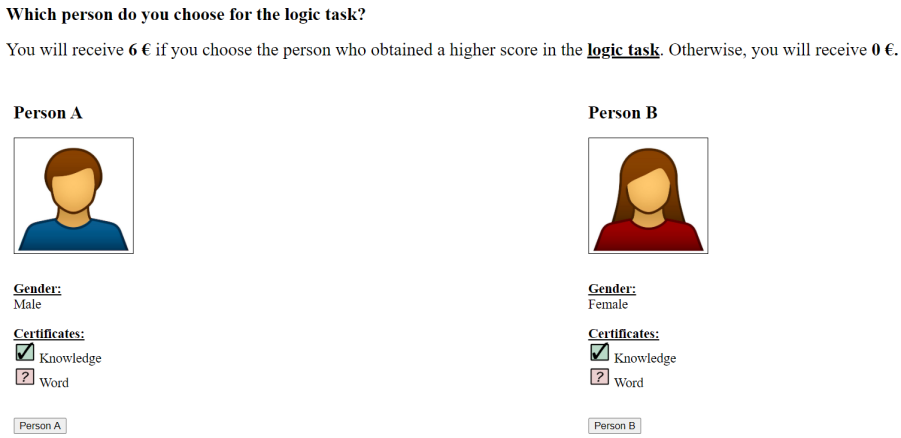
[10]In social psychology, Dovidio and Gaertner (2000) use a methodological approach that is similar in spirit to ours to study aversive racism.

[11]For example, when designing surveys, using the IAT to measure a respondent's implicit biases may be impractical, but adding a few carefully designed (hypothetical) choice questions that vary in how strongly they reveal the decision maker's motives may well be feasible.

experiment, the HIRING EXPERIMENT.[12]

In the HIRING EXPERIMENT, *employers* went through a sequence of nine binary hiring decisions in which they decided which of two job candidates to hire (see Fig. 1 for an example of the screen). In each decision, employers were rewarded when they hired the candidate who performed better in the job task than the other candidate (with ties broken randomly). Across decisions, we systematically varied the CVs of the two candidates. The CV of a candidate included three pieces of information: the gender of the candidate, and information regarding two possible qualifications. This information was provided in the form of a *knowledge certificate* and a *word certificate*, which would certify that a candidate scored in the top 30% (i.e., top 24 out of 80) in the word or knowledge task, respectively. For each of these qualifications, the CV either indicated that the candidate possessed the respective qualification with certainty (the green tick in Figure 1) or that it was unknown whether the candidate possessed that qualification (the question mark in Figure 1). For instance, a female candidate with a knowledge certificate therefore corresponded to a random draw from all female candidates that scored in the top 30% in the knowledge task, irrespective of whether they scored also in the top 30% in the word task or not.

Figure 1: Example screen of an employer's decision setting with the CVs of candidates A and B (order randomized).



Notes: The example reported in the figure is decision 2.

The sequence of nine hiring decisions faced by employers consisted of one *complex* decision (decision 1), where the choice was between a male and female candidate who were *differently qualified*, two gender decisions between *equally qualified* male and female candidates (decisions 2-3), two qualification decisions between *differently qualified* candidates of the same gender (decisions 4-5), and four *simple* decisions, where one candidate was *more*

[12]Appendix A contains specific details regarding the information collected from the job candidates.

qualified (decisions 6-9).[13] Table 1 summarizes the decisions.

A simple hiring decision refers to a choice between a male and a female candidate where one candidate is more qualified, i.e., where one candidate has a certificate whereas the other has none. In the gender and qualification decisions, candidates only differ on one dimension (gender or qualification, respectively). In the complex hiring decision, both candidates are qualified (i.e., have one certificate), but differ on two dimensions (gender and the type of certificate). This is a choice between differently qualified male and female candidates.

The experimental treatments only differed in the complex decision (decision 1). In TREATMENT 1 (T1), the female candidate had a word certificate in the complex decision, and the male candidate had a knowledge certificate. In TREATMENT 2 (T2), the certificates were reversed in the complex decision, with the female candidate possessing a knowledge certificate, and the male candidate having a word certificate.

Table 1: CVs of candidates in all nine hiring decisions

	Candidate A		**Candidate B**		
	Gender	**Certificate**	**Gender**	**Certificate**	
D1 (T1)	F	W	M	K	Complex Choices
D1 (T2)	F	K	M	W	
D2	F	K	M	K	Gender Choices
D3	F	W	M	W	
D4	F	W	F	K	Qualification Choices
D5	M	W	M	K	
D6	F	K	M	?	Simple Choices
D7	F	W	M	?	
D8	F	?	M	K	
D9	F	?	M	W	

Notes: (i) "D1", "D2", etc refer to decision 1, decision 2, etc and "T1" and "T2" refer to treatments 1 and 2. "F" refers to female candidate, "M" refers to male candidate, "W" refers to the word task, while "K" refers to the general knowledge task. (ii) The labels "A" and "B" for the candidates are arbitrary here since they were randomized in the experiment, along with which of the two candidates was presented first on the screen. (iii) The dashed lines indicate blocks of decisions within which the order was randomized between subjects.

In each hiring decision, the employer chose between two of the candidates from the JOB CANDIDATE ASSESSMENT.[14] The employer's incentive was always to hire the candidate that performed better on the job task. Importantly, since the JOB CANDIDATE ASSESSMENT

[13] Our terminology of differently/equally/more *qualified* refers to a gender-blind benchmark, i.e., comparisons where the gender of both candidates is ignored.

[14] More specifically, the employer chose between one candidate randomly selected from the set of candidates that had the same CV as the relevant "A" candidate for that decision, and one candidate randomly selected that had the same CV as the relevant "B" candidate for that decision. The labels "A" and "B" are arbitrary, since they were randomized.

was completed earlier, the employers' hiring decisions had no influence on the candidates' payoffs, nor did the candidates ever learn the employers' decisions. This feature of the design serves two purposes. Firstly, it prevents the hiring decision from influencing the performance on the job task (e.g., similar in spirit to a gift exchange). Secondly, it means we can rule out taste-based discrimination when interpreting our results. To ensure that the employers understood exactly what both qualification tasks and the job task entailed, they were provided with printed sheets which included the questions and problems that job candidates had to solve in the JOB CANDIDATE ASSESSMENT.

Employers earned 6€ if they hired the better candidate in decision 1 (the complex decision) and an additional 6€ if their candidate choice in a randomly drawn decision from 2 to 9 was correct. After each decision, employers could sell their choice for 0.10€. Doing so meant that their hiring choice was replaced by a random draw from the two candidates. We implemented this two-step procedure to gain greater insight into employers' motives. The initial choice forces employers to rank one candidate above the other and thereby reflects hiring decisions in the real world, where one needs to make a concrete choice between distinct options. Therefore, it is our main outcome measure. However, the second step measures employers' desire to actually implement the explicit choice that they made between the candidates. There are at least two reasons why the employer might wish to make use of this option. First, if approximately indifferent, this option allows the employer to express this and earn a small benefit. Second, and more interestingly, it allows an employer to obtain a signaling benefit from initially expressing a preference in line with a social norm, without incurring as large a cost if this is against the employer's actually held belief. The possibility to earn 0.10€ then presents an opportunity to revert such a normative initial choice under the "excuse" or "veil" of a small monetary gain. Only in this latter case would we expect to observe a systematic relationship between the gender of the initial choice and the subsequent purchase of the randomization option.

The order of decisions was partially randomized. Decision 1 was always taken first. It was followed by a block with decisions 2-5, randomly ordered, and then a block with decisions 6-9, randomly ordered. In Table 1, the randomization blocks are separated by dashed lines. This partial randomization was implemented to limit the potential influence of order effects by ensuring that relatively more important decisions for our analysis appeared earlier.

After decision 1, we also measured the employers' beliefs about the relationship between

each of the two certificates and performance in the job task.[15] The reason for this was to assess whether participants shifted their beliefs to justify their decision 1 hiring choices as purely qualification-based. For example, an employer holding a gender bias in favor of the male candidate who happened to have a word certificate in decision 1 might adjust the belief about the informativeness of the word certificate for job performance upwards.

We conducted 10 sessions of the HIRING EXPERIMENT with 24 subjects each. Therefore 240 subjects took part in the experiment as employers, of which N = 119 (49.6%) were female. Mean age was 24.5 years (SD: 4.9 years) and the majority were students in a STEM (50%) or economics/business program (31%). An additional 80 subjects participated in the JOB CANDIDATE ASSESSMENT.[16] The sessions were conducted between October 2017 and January 2018 at the WZB-Technical University laboratory in Berlin. Subjects were invited to participate in the experiment using ORSEE (Greiner, 2015). The experiments were implemented in oTree (Chen, Schonger, and Wickens, 2016), and randomization took place within session, which led to slightly unbalanced treatment assignment (T1: N = 119, T2: N = 121). Demographics by treatment assignment are reported in Appendix B.

2.2 The JOB CANDIDATE ASSESSMENT

The job candidates were sampled in an earlier experiment. This JOB CANDIDATE ASSESSMENT experiment served three purposes. First, it allowed us to run the hiring experiment with candidates drawn from the same pool as the employers, and to populate the candidates' CVs with real qualification data (as opposed to constructing fictitious candidates). This added to the realism of the task and allowed us to incentivize choices, including belief reports. Second, we were thus able to evaluate the decisions of employers against the true distribution of performance in the job task, i.e., to test the accuracy of the employers' revealed beliefs. Third, it provided us with an additional sample of subjects, separate from the employers, from whom we could elicit beliefs about the association of gender with performance in the different tasks, to measure potential stereotypes rampant in the study population.

[15] This was done in the following way. Employers were told that a candidate had been randomly chosen from the pool of candidates and they would earn 3€ if the candidate was in the top 50% in terms of performance on the job task. The employer was given the option to pay to replace this candidate with one who had a word or knowledge certificate, respectively. We elicited subjects' willingness to pay for prices from 0.10€ to 1€ (in steps of 0.10€). 11% of subjects made inconsistent choices in at least one of the lists (i.e., they switched multiple times). After the price list task, subjects were also asked to indicate how informative they thought each of the two certificates was about performance in the job task on two verbal (non-incentivized) 5-point scales from "not informative" to "very informative". This provided a simpler, secondary instrument to measure essentially the same beliefs, given frequent miscomprehension and hence loss of data points in multiple price list tasks (Yu, Zhang, and Zuo, 2020). Our secondary belief measure resembles the common method in closely related psychology research (e.g., Uhlmann and Cohen, 2005).

[16] This comprised 4 sessions of 20 subjects each, of whom 44 were female.

These job candidates completed multiple tasks and were scored on their performance. After the completion of all tasks, one of the tasks was randomly drawn to be paid out, with payoff increasing in their performance. After the tasks, we also elicited job candidates' beliefs about the performance of male and female candidates in the job task. A more detailed description of the tasks and procedures in the JOB CANDIDATE ASSESSMENT experiment, as well as examples of the three tasks are provided in Appendix A.

2.3 Identifying and measuring discrimination

Our design involves several pairs of decisions between CVs in which the only difference is that the gender of the two candidates is reversed. Gender discrimination, conceptualized as the effect of gender information on hiring choices, is identified by any significant difference in the rate at which a given qualification profile is hired over the other one, when it belongs to the male rather than the female candidate. The interpretation of such a difference is therefore that there is relatively more discrimination against one than the other gender. It may be either statistically accurate – when it reflects true underlying differences in the performance of candidates from the two groups – or statistically inaccurate – when it does not reflect true differences in the performance of candidates from the two groups (see, e.g., Bohren et al., 2020).

We define *explicit discrimination* as discrimination when the employer is comparing candidates that are identical other to their gender. This is measured in our decisions 2 and 3: the Gender Choices (see Table 1).

We follow the choice-based definition of *implicit discrimination* proposed by Cunningham and de Quidt (2016), as an implicit preference for one gender that contradicts the explicit preference. It is identified when some employers hire candidates from one gender (e.g., males) in the complex choices, but hire candidates from the other gender (e.g., females) in the explicit gender choices. To conduct this inference here, data from within-subject and between-subject choices needs to be combined. More details on our identification of implicit discrimination are provided along with the results.

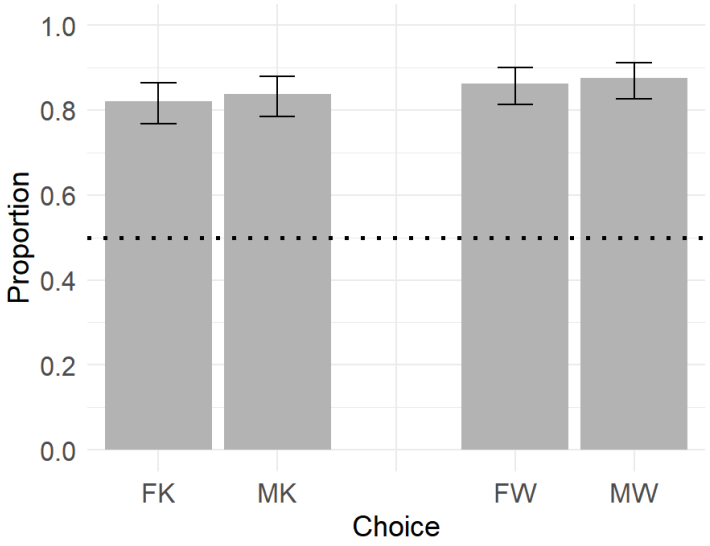
For statistical inference on binomial data we calculate p-values from one-sample (if a proportion is tested against the null hypothesis of 50%) and two-sample (if two proportions are tested against the null hypothesis of zero difference) two-sided score tests (see, e.g., Agresti and Caffo, 2000). For the two-sample case, this is equivalent to a χ^2 test. Correspondingly, the reported confidence intervals are Wilson score intervals.

3 Results

3.1 Comparing hiring decisions in simple and complex choices

We first look at the role of gender in simple and complex hiring decisions.[17] Figure 2 reports the average propensity to choose the more qualified candidate in simple hiring decisions (decisions 6-9 in Table 1). The two left-most bars show hiring propensity when a female or male candidate has a knowledge certificate, and the other candidate (who is always the opposite gender) has no certificate. The two right-most bars analogously show the hiring propensity when a female or male candidate has a word certificate, and the other candidate (who is always the opposite gender) has no certificate. The figure illustrates clearly that in simple hiring decisions, there is no evidence of gender discrimination, with the more qualified candidate being preferred more than 80% of the time, irrespective of whether that candidate is a man or a woman.

Figure 2: Propensity to hire the more qualified candidate in simple hiring decisions.



Notes: (i) In each of the decisions, the alternative choice is a candidate of opposite gender with no certificate (i.e., less qualified). Dashed horizontal line indicates equal propensity (0.5) and 95% confidence intervals are shown; (ii) "F" refers to a female candidate, "M" refers to a male candidate, "K" refers to a candidate having a knowledge certificate, and "W" indicates that the candidate has a word certificate.

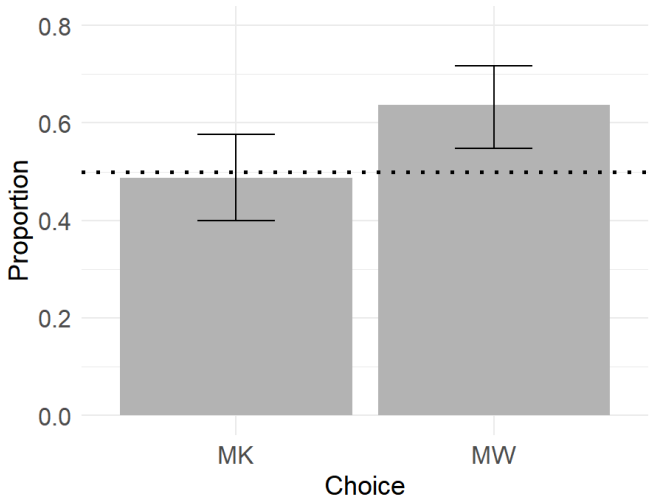
Turning to complex hiring decisions, Figure 3 reports the propensity to hire the male instead of the female candidate when the two candidates hold different certificates (i.e. they are *differently qualified*), making it more opaque which candidate is better qualified. The left bar shows that the male candidate is chosen 48.8% of the time when the male candidate

[17]All choice proportions for all decisions are shown in Appendix B.

has the knowledge certificate, and the female candidate has the word certificate. When the certificates are reversed (i.e., in the other treatment), the male candidate is chosen 63.6% of the time. In the former scenario there is no statistical difference between the rate at which male and female candidates are chosen ($p = 0.78$), while in the latter males are chosen substantially more often ($p = 0.0027$). However, the relevant comparison here is obtained by pooling choices from both treatments such that there are a virtually equal number of hiring choices where the male has the knowledge certificate to where the female has the knowledge certificate.[18] This allows us to estimate the fraction of male candidate choices in complex decisions, while maintaining gender-symmetry in terms of qualification certificates. Doing this, we see that overall males are chosen 56.3% of the time, which is greater than the no-discrimination benchmark of 50% ($p = 0.053$). This shows that when hiring decisions become more complex, we see male-favoring discrimination amongst the same group of subjects who displayed no discrimination in the simple hiring decisions.

Arguably, complex decisions increase subjectivity in judgment compared to simple decisions. This is broadly in line with the randomized field experiment by Bohren et al. (2019), who find "greater discrimination against females when judgments of quality are more subjective", as well as with results from the psychology literature: For example, Dovidio and Gaertner (2000) report more racial discrimination in simulated hiring decisions when candidate information is more ambiguous with respect to qualification.

Figure 3: Propensity to hire a male candidate in the complex hiring decisions



Notes: (i) In each of the decisions, the alternative choice is a female candidate who has the other certificate (i.e., who is differently qualified). Dashed horizontal line indicates equal propensity (0.5), and 95% confidence intervals are shown.

[18]Treatment 1 has slightly fewer subjects (N = 119) than Treatment 2 (N = 121). This is due to conducting the within-session randomization at the individual level.

The discussion so far raises several questions. Firstly, a natural question is whether the discriminatory hiring observed can be explained purely by accurate statistical reasoning? This might be the case if men perform better than women, implying that the signal that a candidate is a man contains information about their expected performance. We will assess this by studying the actual performance observed by men and women in the task that subjects are hired for. Then, we evaluate the degree to which we find explicit discrimination versus implicit discrimination in the hiring data.

3.2 Do men actually perform better than women?

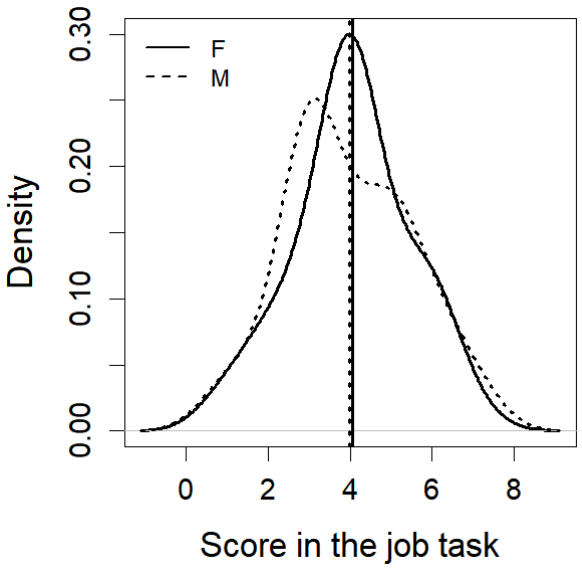
In order to address the question of whether the higher propensity to hire male candidates is consistent with accurate statistical discrimination, we look at the distribution of candidates' performance in the job task. Descriptive performance statistics by gender are shown in Table 2, and the smoothed distribution of scores is shown in Figure 4. There is no statistical difference between the mean score for men and the mean score for women (two-sided two-sample t-test: $p = 0.82$); moreover, all three quartiles are identical.

Table 2: Descriptive statistics on actual performance in the job task by gender.

Statistic	Female	Male
N	44	36
Mean	4.05	3.97
SD	1.38	1.52
Min	1	1
25% Pctl	3	3
Median	4	4
75% Pctl	5	5
Max	7	7

At face value, this suggests that there is no statistical justification for the preferential hiring of men. However, it is worth noting that since subjects are paid according to whether the candidate they hired performed better, the relevant statistic is rather the probability that a randomly drawn male performs better than a randomly drawn female.

Figure 4: Kernel density of job task performance by gender



Notes: (i) The solid and dashed lines report scores of female (F) and male (M) candidates in the job task (number of correctly solved matrix logic exercises), (ii) Vertical lines show group means.

Overall, the probability that a randomly drawn man (woman) performed strictly better than a randomly drawn woman (man) was 39.1% (44.8%). In 16.1% of all paired comparisons, both performed equally well. This indicates that when considering all the candidates in the job assessment task pool, female candidates were slightly more likely to have performed better than male candidates overall. Restricting attention to the relevant subgroups of candidates being compared in each of the 9 decisions, Table 3 reports the conditional probabilities that a randomly drawn candidate with the characteristics of Candidate A performed better than a randomly drawn candidate with the characteristics of Candidate B (and vice versa). It shows who should be chosen in each of the decisions according to the true probability of having a higher score in the job task. In half of the decisions in which the employer had to choose between a male and female candidate, selecting the female candidate maximized expected earnings. This leads to the conclusion that, in our data, there is no empirical basis for beliefs that overall favor men over women as better employees for the job task. By contrast, a more qualified person, as in decisions 5–9, should indeed always be hired, confirming the value of both certificates irrespective of gender.[19]

[19]When the top 30% from the knowledge and word task were selected into the pool of certificate holders, ties were broken randomly. Out of the 24 candidates with a knowledge [word] certificate, 10 [13] were women.

Table 3: True probabilities of performing strictly better in the job task for all comparisons of candidates in the hiring decisions. Probabilities per row do not add up to 1 due to ties.

	Candidate A	Candidate B	Prob(A>B)	Prob(B>A)
D1 (T1)	FW	MK	0.385	0.505
D1 (T2)	FK	MW	0.491	0.345
D2	FK	MK	0.521	0.314
D3	FW	MW	0.350	0.538
D4	FW	FK	0.154	0.531
D5	MW	MK	0.429	0.383
D6	FK	M?	0.594	0.247
D7	FW	M?	0.432	0.429
D8	F?	MK	0.396	0.461
D9	F?	MW	0.366	0.492

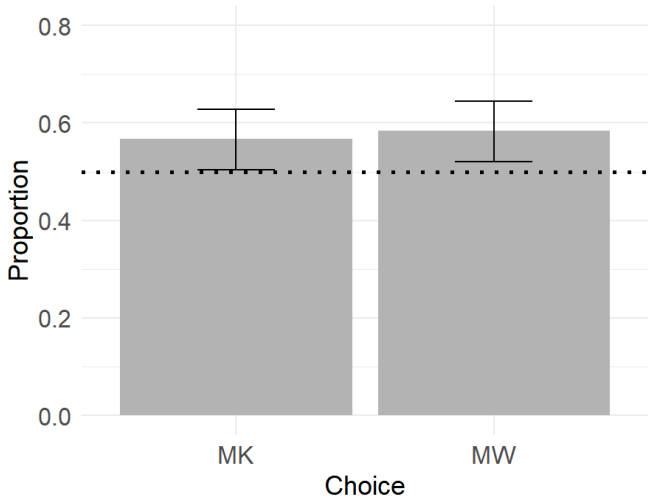
3.3 Explicit Discrimination

The discussion above has illustrated that: (i) participants hire the more qualified candidate most of the time when it is clear who is more qualified, irrespective of gender, (ii) participants hire males preferentially in complex decisions, where it is unclear which candidate is better qualified; (iii) this preferential hiring of males is not statistically justified according to the objective performance of males and females in the job task of interest.

A question that therefore remains is what is driving the observed preferential hiring of males over females and in which settings. Our design and the fact that there is no gender difference in actual job performance renders classical taste-based and (rational/accurate) statistical discrimination ineffective in explaining the results. A remaining potential explanation is that participants (falsely) believe that men outperform women in the job task, what Bohren et al. (2020) refer to as *inaccurate statistical discrimination*.

To assess whether participants in our experiment tend to display *explicit* inaccurate statistical discrimination, we consider hiring decisions where the only characteristic that differs between the two candidates is their gender (decisions 2-3). More specifically, we ask whether males and females are hired equally often when both candidates have the same certificate. Figure 5 shows that when both candidates have a knowledge certificate, males are hired in 56.7% of cases, and when both candidates have a word certificate, males are hired in 58.3% of cases. Both proportions are statistically significantly different from an equal hiring of men and women ($p = 0.039$ and $p = 0.0098$, respectively). Thus, as in Coffman et al. (2020), even when CVs are otherwise identical and it is salient that a choice is only about gender, there is a male bias in hiring.

Figure 5: Explicit gender discrimination in hiring decisions between equally qualified candidates.



Notes: (i) The figure reports the propensity to hire a male candidate in decisions where the alternative choice is a female candidate who has the same certificate (i.e., who is equally qualified), (ii) Dashed horizontal line indicates equal propensity (0.5), and 95% confidence intervals are shown.

3.4 Implicit Discrimination

Having established the presence of explicit inaccurate statistical discrimination, we now turn to our novel test for the *additional* presence of what we call implicit discrimination (see Cunningham and de Quidt, 2016) in the population.

Such discrimination would manifest itself here only in situations where it is not possible to establish unambiguously that one candidate is more qualified than another, i.e., in complex hiring decisions where candidates have different qualifications. Implicit discrimination can particularly be expected if such a situation is paired with a pervasive stereotype within society that favors one group in the relevant job domain, but a conflicting social stigma or institutional rule against the preferential treatment of members of that group over another.

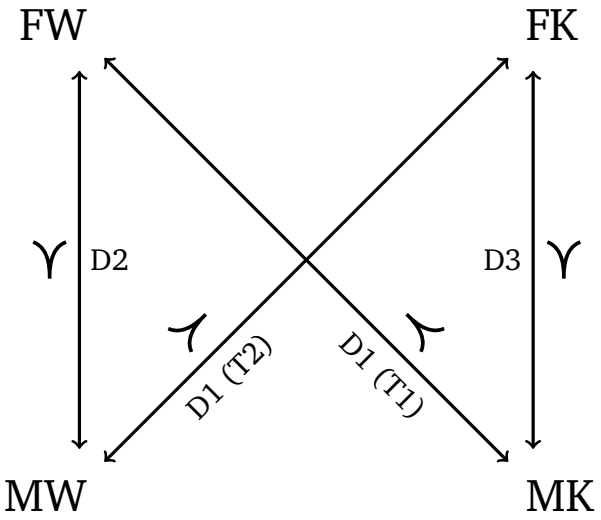
In a context of this nature, while simple decisions with one candidate more qualified than the other might depend on merit considerations only, beliefs about gender (in particular, inaccurate stereotypes, which we consider below in section 3.6) would rise to behavioral significance in complex ones.

Our results presented so far already provide suggestive evidence in favor of the presence of implicit discrimination. In simple decisions, qualification considerations seem to be viewed as more informative than gender considerations (Figure 2), and yet, complex hiring decisions that trade off gender and qualification differences still produce an overall bias towards males (Figure 3). We now apply the identification framework proposed by

Cunningham and de Quidt (2016) to our setting to test for an implicit bias against females as a systematic violation of rational choice, by combining the between-subjects choice data from complex decision 1 with the within-subjects choice data from gender decisions 2-3.

Consider the pattern of binary choices presented in Figure 6, all of which are part of our study, recalling that choosing person *A* over person *B* here means "believing that *A* performs better in the target task than *B*." We interpret here any choice as a strict hiring preference.

Figure 6: The choice pattern of implicit discrimination against females



This pattern of behavior is generated when an employer chooses the female whenever a male and female candidate have the same certificate, but chooses the male whenever a male and female candidate hold different certificates. If an employer's choices follow this pattern, this cannot be rationalized by any single preference (more specifically, belief) defined over these options, because they imply cycles: MK $\prec$ FK $\prec$ MW $\prec$ FW $\prec$ MK.[20] Cunningham and de Quidt (2016) interpret this particular pattern as an implicit preference for males, and they discuss several motives and mechanisms that could give rise to this.

The pattern reveals two conflicting preferences/beliefs, with the particular choice set available determining which one is revealed. Intuitively, the vertical choices (decisions 2-3) reveal the belief that females perform better while the diagonal choices (decision 1) reveal the belief that males perform better. This behavior can arise when one holds a prejudicial belief against females that is not revealed explicitly in any choice that is solely about gender, and it is clearly inconsistent with always choosing the candidate who is believed to be

[20]The coincidence in terminology here between beliefs and preferences is due to the fact that one's preference ordering over candidates should be a function of one's beliefs about the mapping from different candidate resumes to performance in the job task.

better.[21]

We designed our study with a *between-subjects* design for the diagonal choices (each corresponds to decision 1 of one treatment) to avoid order effects. The vertical choices are made by every subject, so we can use a *within-subject* analysis. For the choice data generated by our design, we now derive a lower bound on the fraction of participants that discriminate *only* implicitly.

Consider the sub-population of employers E that choose the female candidate in both decisions 2 and 3. We are interested in the fraction $\sigma_E(MK, MW)$ of such employers, who prefer the male candidate in both of the two different versions of decision 1, i.e., who prefer MK over FW (decision 1 of treatment 1) as well as MW over FK (decision 1 of treatment 2). As we show in Appendix C, under the assumption that each of these two preferences is treatment-independent, this fraction $\sigma_E(MK, MW)$ is bounded from below by

$$l_E^M = \max\left\{0, \sigma_{1,E}^{MK} - \sigma_{2,E}^{FK}\right\},$$

where $\sigma_{1,E}^{MK}$ is the fraction of employers in E choosing MK over FW, as observed from decision 1 of treatment 1, and $\sigma_{2,E}^{FK}$ is the fraction of employers in E choosing FK over MW, as observed in decision 1 of treatment 2. (Since the binary choice fractions add up to one within every treatment, the lower bound can equivalently be written as $l_E^M = \max\left\{0, \sigma_{2,E}^{MW} - \sigma_{1,E}^{FW}\right\}$.) The difference in the lower bound expression holds the certificate constant and compares which gender was chosen more often, but—in contrast to decisions 2 and 3—it does so indirectly, by comparing between the treatments.

As defined, implicit discrimination concerns only employers that do not discriminate explicitly already. This means that it is an *additional* form of discrimination. Since we have already observed a substantial amount of explicit discrimination in favor of males, the restriction to employers that choose the female in both decisions 2 and 3 (equally qualified candidates of different gender) implies a fairly small sub-population E: Approximately one out of every five employers enter this sub-sample (45 out of 240); by comparison, a third of all employers choose the male in both of those decisions (81 out of 240). Nonetheless, we find evidence of implicit discrimination within our sample. Since we identify a lower bound, it is probable that we are underestimating the true magnitude. Applying the logic described above, the first row of Table 4 provides the calculations for detecting implicit discrimination against women in our sample. It shows that even amongst the subset of employers who choose the female candidate in both cases where the candidates were *equally qualified* (sub-population E), when the candidates where *differently qualified*, they chose: (i) a male

[21] Our study differs from complementary work by Bohnet et al. (2015) because they do not identify implicit discrimination in the way it is defined here. One reason for this is that they do not elicit choices where both genders are equally qualified (i.e., the verticals in Figure 6).

candidate with a knowledge certificate (46%) more often than a female candidate with a knowledge certificate (35%), and (ii) a male candidate with a word certificate (65%) more often than a female candidate with a word certificate (54%).[22] These choice fractions imply that we can place a lower bound on the implicit discrimination in favor of men observed in sub-population E of approximately 11% and, consequently, in the whole population of approximately 2% (95% CI: 0.1%-4.8%).

Table 4: Lower bound of implicit discrimination in hiring decisions

Against	Sub-population E defined by	$\sigma_{1,E}^{MK}$	$\sigma_{2,E}^{FK}$	$\sigma_{2,E}^{MW}$	$\sigma_{1,E}^{FW}$	Lower bound in E	Lower bound in full sample
F	choose F in D2 & D3 (N=45)	0.46	0.35	0.65	0.54	0.111	0.021
M	choose M in D2 & D3 (N=81)	0.53	0.33	0.67	0.47	0	0
W	choose W in D4 & D5 (N=76)	0.27	0.15	0.85	0.73	0	0
K	choose K in D4 & D5 (N=126)	0.80	0.74	0.26	0.20	0	0

As a placebo check, we carry out identical calculations to check for implicit discrimination on any other dimension observed in the candidates' CVs: (i) being male, (ii) holding a knowledge certificate and (iii) holding a word certificate. The bottom three rows of Table 4 report these calculations (note that the sub-sample E is redefined in each row). In all three cases, we find that the lower bound on implicit discrimination against that characteristic is zero. This provides us with confidence that the implicit discrimination against females we observe is systematic.

3.5 Assessing evidence for signaling motives among employers

How can this observed implicit discrimination be explained? One potential foundation is image concerns implying signaling motives, as shown by Cunningham and de Quidt (2016):[23] In contrast to the decisions where candidates only differ in gender (decisions 2-3), the hiring choice in the complex decision (decision 1) cannot be unambiguously related to a gender bias and is thus less likely to harm the social or self image of the employer (as in theories of identity management, Bénabou and Tirole, 2011). We can use two additional design features of our experiment to assess the evidence for the role of signaling/image concerns in our population of employers more broadly. In particular, we can ask: Do we see behavior that is consistent with attempts to cover up discrimination?

[22]To be precise, this refers to the decisions made by employers in decision 1, where they compare a male candidate with a knowledge [word] certificate to a female candidate with a word [knowledge] certificate in treatment 1 [2]. Therefore, these comparisons are between-subject comparisons.

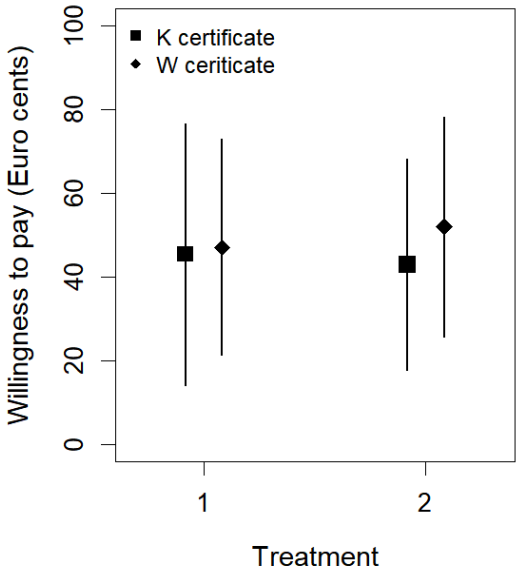
[23]In the closely related social psychology literature discussing aversive racism, signaling motives are also discussed as playing a key role (Hodson et al., 2010).

First, after having taken the complex decision (decision 1), employers' beliefs about the correlation between a qualification – a knowledge certificate, or a word certificate – and performance in the job task were elicited in two multiple price lists. This incentivized employers to indicate which of the two qualifications they consider more informative for job performance. Crucially, this allows them to justify gender discrimination: any gender-driven choice could then have been rationalized as being qualification-driven (by adjusting the relative importance of the two qualifications). This has been called "redefining merit" by psychologists Uhlmann and Cohen (2005), who describe shifting preferences for qualifications as a way to "allow people to maintain an image of themselves as objective and principled" (p. 479) despite discriminating by gender when hiring employees. Since the treatments differed by who (the male or the female candidate) had which certificate in the complex decision, we can look at whether there are treatment differences in the employers' post hoc willingness to pay for each of the two certificates (Figure 7). Estimating the relevant difference in differences with robust standard errors clustered at the employer level reveals a small, non-significant tendency for employers to put a higher price on the word certificate relative to the knowledge certificate in the treatment where it is the male candidate having the word certificate (0.07 €, 95%-CI: -0.02–0.17 €, p = 0.14). If we use the simpler belief elicitation method to impute the data points missing due to multiple switching in the price lists (42 out of 480 responses),[24] we gain statistical power and more certainty about this effect (0.08 €, 95%-CI: -0.01–0.17 €, p = 0.079; Appendix B contains full regression results). It is worthwhile noting that we would expect a causal effect only on those employers that have an interest in covering up their motives, which may explain the rather small estimated effect size.

Second, after each of the nine hiring decisions in our experiment, employers had the opportunity to earn 0.10€ by replacing their explicit candidate choice with a random draw between the two candidates. However, this amount was negligibly small in comparison to the incentives on offer for hiring decision the correct candidate (6€). Employers with only financial motives should thus only sell their choice if their belief about who is the better candidate is extremely weak. An alternative reason why employers might exercise this option of selling their hiring choice is that it allows them to switch back to a 50:50 chance of hiring either of the candidates *after* having signalled a preference for one of the candidates. It therefore reduces the cost of sending the socially desirable signal (through their initial decision), with the 0.10€ then presenting an "excuse" or "veil" for switching away from their initial decision (similar to the idea that adding small amounts of risk to the

[24] We exploit the strong theoretical and empirical correlation between the responses in the belief elicitation via multiple price lists and via non-incentivized ordinal 5-point-scales ($r = 0.48$, $p < 0.001$) and predict the missing values in willingness to pay via OLS regression.

Figure 7: Beliefs about qualification informativeness



Notes: (i) The figure reports the elicited willingness to pay of employers for a candidate that has a knowledge (K) [word (W)] certificate as replacement for a randomly drawn candidate, (ii) means and standard deviations are shown.

choice setting could present an excuse for gender discrimination, see Coffman et al., 2020). Indeed, we find that employers are more likely to sell their choice after hiring a female (23%) than after hiring a male candidate (11%, p = 0.014) in the complex decision. The same holds true in the aggregate across all the seven decisions in which candidates differ by gender (32% vs. 27%, p = 0.013).

3.6 Beliefs about gender in the candidate sample

Multiple strands of evidence presented in our results demonstrate that employers (explicitly or implicitly) discriminate against female candidates. To substantiate the claim that biased beliefs about the superiority of men (i.e., stereotypes) in this domain were responsible for the observed behavior, it is instructive to elicit the beliefs about the association between gender and performance in the job task more directly. We did this among subjects in the Job Candidate Assessment, i.e., in a sample from the same university student subject pool as the employers that was familiar with the tasks, but had not taken part in the Hiring Experiment.

On average, job candidates believed that in 55.4% (SD: 16.2%) of comparisons between a randomly drawn man and a randomly drawn woman, the former performed better in the job task. This is statistically greater than 50% (one-sample t-test: p = 0.0037) and provides evidence of a more generally held (inaccurate) belief in the population that men perform

better in the logic task.

4 Conclusion

The economics literature has typically dichotomized discrimination into taste-based (Becker, 1957) and (accurate) statistical discrimination (Phelps, 1972; Arrow, 1973). Even though our experimental design (i) does not permit employers to preferentially reward or choose to interact with candidates from one gender, and (ii) involves a job task where there is no gender gap in performance we still observe substantial discrimination against women in a hiring task. In particular, we observe both explicit and implicit belief-based discrimination consistent with a statistically inaccurate gender stereotype that women perform worse in a logic task. Further, our results highlight the importance of the choice setting for determining whether and how discrimination will manifest.

One important caveat to our results is that the degree of implicit discrimination that we detect is likely to be underestimated in relation to its occurrence in natural settings in the general population. There are several reasons for this. First, the experimental design only permits us to place a lower bound on the degree of implicit discrimination observed in our experiment. Second, our population is comprised of young and highly educated students who are likely to hold less gender-stereotyped beliefs than the general population.[25] Third, the hiring decisions that subjects make in our experiment are anonymous. Therefore, the role of social image concerns is substantially dampened. Since implicit discrimination involves a tension between an underlying preference and the signal that one's actions send (to oneself and others), the dampening of social image concerns is likely to yield a shift towards more explicit discrimination and less implicit discrimination. In many real-world contexts, decisions are not anonymous, and one would expect that the increased role of social image would lead to a shift away from explicit discrimination. At the same time, real-world hiring decisions are typically complex, especially among the "finalists" in a hiring process. In such scenarios, implicit discrimination is likely to play a larger role. Together, these considerations point towards the worrying conclusion that if we are able to detect implicit discrimination in the stark, anonymous environment of our experiment, it is likely to be substantially more prevalent in real world contexts.

A key question to address in future research, therefore, is: What are the contextual and institutional factors that are likely to generate implicit discrimination? As implicit discrimination can result from a conflict between what an individual would like to do (preferences),

[25] A large fraction of the participants in our experiment attend a technical university, implying that they interact regularly with male and female classmates that are selected to be above-average in terms of their quantitative abilities. This may serve to ameliorate gender stereotypes they previously held.

and what is socially acceptable behavior (norms),[26] it follows that it is more likely to be observed in hiring scenarios with the following characteristics. Scenarios where: (i) preferentially hiring a candidate from a particular group is socially stigmatized, (ii) many individuals in the population of decision makers hold stereotypes (or tastes) that favor this group, (iii) information about the hiring process can be publicly observed, (iv) the job candidates are (horizontally) heterogeneous, or the expertise and attributes required for the job are more opaque (i.e., the "revealingness" of the hiring decisions about biases is low).

One important lesson from the recent discrimination literature is that it is imperative that policy interventions are tailored to address the source of the problem. In the case of implicit discrimination, the design of policy interventions depends critically on whether individuals are masking their discriminatory preferences from themselves (self-image) or others (social image) – i.e., whether they are aware of their bias or not. In situations where individuals are unaware of their own bias, it may be sufficient to inform these individuals about the bias present in their own or other individual's past decision making. Alesina et al. (2018) demonstrate that this can be effective in de-biasing teachers with implicit discriminatory preferences. If instead, individuals are aware of their bias and are hiding their preferences from others, the policy prescription is very different. Here, carefully designed procedures, such as requiring clear and transparent ex ante decision rules that leave little wiggle room might be more effective (see, e.g., Uhlmann and Cohen, 2005).

In cases where inaccurate gender-biased beliefs or stereotypes are at the heart of discrimination, as presented here, confronting these beliefs with information can also be an effective approach. This solution is discussed by Bohren et al. (2020). Bordalo et al. (2016) argue that stereotypes are typically based on a "kernel of truth".[27] If one can demonstrate in a particular context that there are no statistical differences between two groups, this may induce a re-evaluation of the stereotype. However, discriminatory beliefs can be sticky even in the presence of informative signals that contradict them (Reuben et al., 2014). This may especially be the case when a motivation exists to maintain false beliefs against incoming data (as for favorable in-group beliefs, demonstrated by Cacault and Grieder, 2019). Such motivated tastes over beliefs are harder to combat – doing so requires influencing the formation of preferences, which is a complex process taking place over a long period of time

[26] In situations with these characteristics, the motive to discriminate explicitly is reduced by social stigma. For example, Barr, Lane, and Nosenzo (2018) provide evidence that discrimination is reduced when it is perceived to be more socially inappropriate (although, they focus on taste-based discrimination). We argue here that depending on the context, these underlying preferences may instead manifest as implicit instead of explicit discrimination.

[27] However, it is important to note that the "kernel of truth" may be the result of endogenous processes in society that make stereotypes self-fulfilling. For example, Chauvin (2018) demonstrates that in a society where individuals are prone to exhibit the Fundamental Attribution Error, they underestimate the role played by differing circumstances on the outcomes of different groups, and therefore form biased beliefs about underlying characteristics of these groups.

and not easy to influence. Lai et al. (2016) show that brief interventions like presenting counter-stereotypical examples are unlikely to have long-lasting impacts on implicit bias. Further, Dovidio et al. (2016) discuss how many well-intentioned interventions aimed at reducing intergroup bias may backfire.

Interestingly, our results imply that hiring procedures which force joint rather than separate evaluation of candidates, as suggested by the lab experiments of Bohnet et al. (2015), are not a panacea when performance signals are less straightforward to interpret (i.e., there is not a clear and simple correspondence between qualifications and the job being hired for) and do not allow one to unambiguously rank one candidate over the other. In line with their results though, we find no gender discrimination in joint evaluations of female-male candidate pairs where ranking by qualification is simple (in our case, one certificate vs. no certificate).

Thus, together with the contemporary discrimination literature, this paper highlights that in order to find effective remedies to combat discrimination, it is crucial to have a fine-grained and accurate understanding of the underlying causes of discrimination and to be able to detect the different manifestations that discriminatory preferences can take in different contexts. The paper also demonstrates a central role for beliefs in the formation of discriminatory behavior.

Future work in this area might investigate the relative importance of self-image and social-image in generating implicit discrimination, and systematically study the contextual and institutional factors that exacerbate and alleviate it. Lessons learned from these exercises would be invaluable for designing effective policy tools that are able to treat the underlying problem, as opposed to just treating the symptoms and allowing discrimination to simply manifest in a different form.

References

Agresti, A. and B. Caffo (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician 54*(4), 280–288.

Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2018). Revealing stereotypes: Evidence from immigrants in schools. *Working Paper*.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter and A. Rees (Eds.), *Discrimination in Labor Markets*. Princeton University Press.

Ayres, I. and P. Siegelman (1995). Race and gender discrimination in bargaining for a new car. *American Economic Review 85*(3), 304–321.

Banaji, M. R. and A. G. Greenwald (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology 68*(2), 181.

Barr, A., T. Lane, and D. Nosenzo (2018). On the social inappropriateness of discrimination. *Journal of Public Economics 164*, 153–164.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics 126*(2), 805–855.

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

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. In A. Banerjee and E. Duflo (Eds.), *Handbook of Field Experiments*. North Holland.

Bertrand, M. and S. Mullainathan (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.

Biernat, M. and D. Kobrynowicz (1997). Gender-and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology 72*(3), 544.

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

Bohnet, I., A. Van Geen, and M. Bazerman (2015). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science 62*(5), 1225–1234.

Bohren, A., K. Haggag, A. Imas, and D. G. Pope (2020). Inaccurate statistical discrimination. *Working Paper*.

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

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

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

Bowles, H. R., L. Babcock, and L. Lai (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes 103*(1), 84–103.

Cacault, M. P. and M. Grieder (2019). How group identification distorts beliefs. *Journal of Economic Behavior & Organization 164*, 63 – 76.

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2020). Are referees and editors in economics gender neutral? *Quarterly Journal of Economics 135*(1), 269–327.

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

Charles, K. K. and J. Guryan (2011). Studying discrimination: Fundamental challenges and recent progress. *Annual Review of Economics 3*(1), 479–511.

Chauvin, K. P. (2018). A misattribution theory of discrimination. Technical report, Mimeo, December.

Chen, D. L., M. Schonger, and C. Wickens (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Coffman, K. B., C. L. Exley, and M. Niederle (2020). The role of beliefs in driving gender discrimination. *Management Science (forthcoming)*.

Coffman, K. B., C. B. Flikkema, and O. Shurchkov (2019). Gender stereotypes in deliberation and team decisions. *Working Paper*.

Corno, L., E. La Ferrara, and J. Burns (2019). Interaction, stereotypes and performance: Evidence from South Africa. *IFS Working Papers*.

Cunningham, T. and J. de Quidt (2016). Implicit preferences inferred from choice. Mimeo.

Danilov, A. and S. Saccardo (2017). Discrimination in disguise. Mimeo.

Dovidio, J. F. and S. L. Gaertner (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science 11*(4), 315–319.

Dovidio, J. F., S. L. Gaertner, E. G. Ufkes, T. Saguy, and A. R. Pearson (2016). Included but invisible? Subtle bias, common identity, and the darker side of "we". *Social Issues and Policy Review 10*(1), 6–46.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies 83*(2), 587–628.

Glick, P., C. Zion, and C. Nelson (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology 55*(2), 178.

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

Goldin, C. and C. Rouse (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review 90*(4), 715–741.

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

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association 1*(1), 114–125.

Hengel, E. (2020). Publishing while female. Are women held to higher standards? Evidence from peer review. *Working Paper*.

Hilton, J. L. and W. Von Hippel (1996). Stereotypes. *Annual Review of Psychology 47*(1), 237–271.

Hodson, G., J. F. Dovidio, and S. L. Gaertner (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin 28*(4), 460–471.

Hodson, G., J. F. Dovidio, and S. L. Gaertner (2010). The aversive form of racism. In J. L. Chin (Ed.), *Race and ethnicity in psychology. The psychology of prejudice and discrimination: Racism in America*, Volume 1, pp. 119–135. Praeger Publishers.

Isaksson, S. (2018). It takes two: Gender differences in group work. *Working Paper*.

Jowell, R. and P. Prescott-Clarke (1970). Racial discrimination and white-collar workers in britain. *Race 11*(4), 397–417.

Judd, C. M. and B. Park (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review 100*(1), 109.

Kübler, D., J. Schmid, and R. Stüber (2018). Gender discrimination in hiring across occupations: a nationally-representative vignette study. *Labour Economics 55*, 215–229.

Kurdi, B., A. E. Seitchik, J. R. Axt, T. J. Carroll, A. Karapetyan, N. Kaushik, D. Tomezsko, A. G. Greenwald, and M. R. Banaji (2019). Relationship between the implicit association test and intergroup behavior: A meta-analysis. *American psychologist 74*(5), 569.

Lai, C. K., A. L. Skinner, E. Cooley, S. Murrar, M. Brauer, T. Devos, J. Calanchini, Y. J. Xiao, C. Pedram, C. K. Marshburn, S. Simon, J. C. Blanchar, J. A. Joy-Gaba, J. Conway, L. Redford, R. A. Klein, G. Roussos, F. M. H. Schellhaas, M. Burns, X. Hu, M. C. McLean, J. R. Axt, S. Asgari, K. Schmidt, R. Rubinstein, M. Marini, S. Rubichi, J.-E. L. Shin, and B. A. Nosek (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General 145*(8).

Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review 90*, 375–402.

McIntyre, S., D. J. Moberg, and B. Z. Posner (1980). Preferential treatment in preselection decisions according to sex and race. *Academy of Management Journal 23*(4), 738–749.

Milkman, K. L., M. Akinola, and D. Chugh (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology 100*(6), 1678.

Neumark, D., R. J. Bank, and K. D. Van Nort (1996). Sex discrimination in restaurant hiring: An audit study. *Quarterly Journal of Economics 111*(3), 915–941.

Newman, J. M. (1978). Discrimination in recruitment: An empirical analysis. *Industrial and Labor Relations Review 32*(1), 15–23.

Oswald, F. L., G. Mitchell, H. Blanton, J. Jaccard, and P. E. Tetlock (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology 108*(4), 562–571.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review 62*(4), 659–661.

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

Riach, P. A. and J. Rich (1987). Testing for sexual discrimination in the labour market. *Australian Economic Papers 26*(49), 165–178.

Riach, P. A. and J. Rich (2002). Field experiments of discrimination in the market place. *The Economic Journal 112*(483), F480–F518.

Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics 17*(3), 523–534.

Sarsons, H. (2019). Interpreting signals in the labor market: Evidence from medical referrals. *Working Paper*.

Sarsons, H., K. Gërxhani, E. Reuben, and A. Schram (2020). Gender differences in recognition for group work. *Working Paper*.

Small, D. A., M. Gelfand, L. Babcock, and H. Gettman (2007). Who goes to the bargaining table? The influence of gender and framing on the initiation of negotiation. *Journal of Personality and Social Psychology 93*(4), 600.

Snyder, M. L., R. E. Kleck, A. Strenta, and S. J. Mentzer (1979). Avoidance of the handicapped: An attributional ambiguity analysis. *Journal of Personality and Social Psychology 37*(12), 2297–2306.

Uhlmann, E. L. and G. L. Cohen (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science 16*(6), 474–480.

Yinger, J. (1986). Measuring racial discrimination with fair housing audits: Caught in the act. *The American Economic Review*, 881–893.

Yu, C. W., Y. J. Zhang, and S. X. Zuo (2020). Multiple switching and data quality in the multiple price list. *The Review of Economics and Statistics (forthcoming)*.

A The Job Candidate Assessment: Tasks and Procedure

A.1 The word task (word certificate)

Subjects solved three word search puzzles.[28] They had 90 seconds to work on each puzzle. Each of the puzzles contained 10 hidden words, and subjects were presented with 30 possible answers. Subjects selected answers they thought were correct. For each correct answer, subjects gained 0.40€. For each wrong answer, they lost 0.40€ (we restricted the total payoff in this task to be non-negative). On average, subjects earned 5.03€ (SD: 2.69€) in this task. Performance was measured as the total number of selected correct response options minus the number of selected incorrect response options.

A.2 The knowledge task (knowledge certificate)

This task consisted of 30 general knowledge questions, for which four minutes were available. Questions were selected from several categories (geography, environmental sciences, pop culture, arts, literature and history) but were presented in an arbitrary order. Four response options were presented for each question, of which only one was correct. Each correct answer was worth 0.60€. On average, subjects earned 5.08€ (SD: 2.14€) in this task. Performance was measured as total number of questions answered correctly.

A.3 The logic task (job task)

Subjects solved matrix reasoning exercises of the type that are commonly used in IQ tests. Each of the ten questions consisted of a 3-by-3 matrix in which one cell was empty. Matrices had to be completed by choosing one of the six response options.[29] Subjects were given five minutes to work on this task. They earned 1.30€ for each matrix problem they solved correctly. On average, they earned 5.22€ (SD: 1.87€) in this task. Performance was measured as total number of matrix exercises solved correctly.

A.4 Procedure in the Job Candidate Assessment

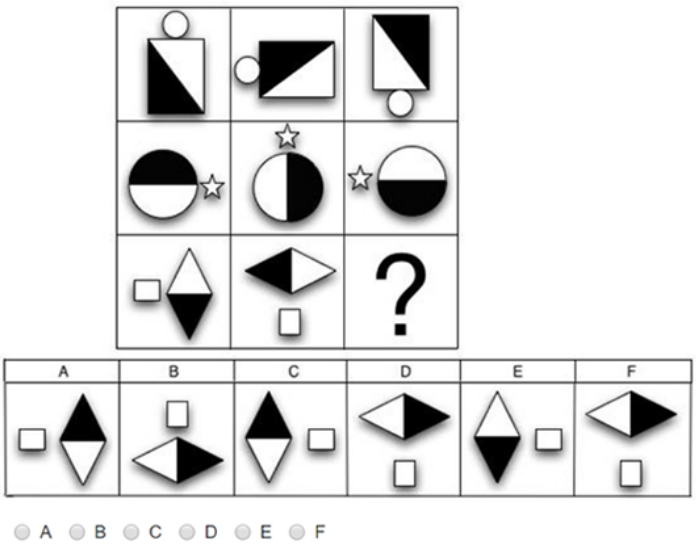
The order of the tasks was held constant across all subjects. After the completion of all tasks, an incentivized belief elicitation followed. For each of the tasks mentioned above, subjects were asked how often they believed a randomly drawn male would perform better than

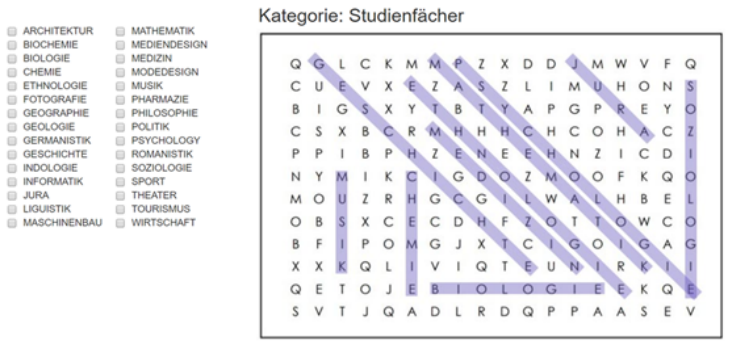
[28] Each puzzle had a theme: animals, countries or fruit.

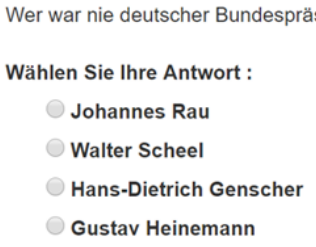
[29] Matrix exercises were taken from the online resources of the ICAR project. See: https://icar-project.com/projects/icar-project

a randomly drawn female.[30] Subjects' beliefs were incentivized by means of a quadratic scoring rule.

Figure 8: Examples of the logic task (upper panel), the knowledge task (middle panel) and the word task (lower panel). These examples and their solutions were shown to subjects in the job candidate assessment as part of the instructions before they worked on the problems they were scored for.







[30]More specifically, we asked subjects to think about taking 100 draws of a pair of subjects, each containing a randomly drawn male and a randomly drawn female from their session. They were asked to indicate how often they believed that the randomly drawn male performed better than the randomly drawn female in the respective task (with ties broken randomly).

B Additional Tables and Figures

Table 5: Subject demographic information from the HIRING EXPERIMENT

Variable	Treatment 1	Treatment 2
Age (mean, SD)	24.8 (5.4)	24.2 (4.4)
Gender: female (N, %)	59 (49.6)	60 (49.6)
Study subject: STEM (N, %)	56 (50)	65 (53.7)
Study subject: Economics/Business (N, %)	37 (33)	38 (31.4)

Figure 9: Choice data for each of the nine hiring decisions; for all subjects (upper panel) and separated by treatment (middle and lower panel). For explanation of decision numbers, see Table 1.

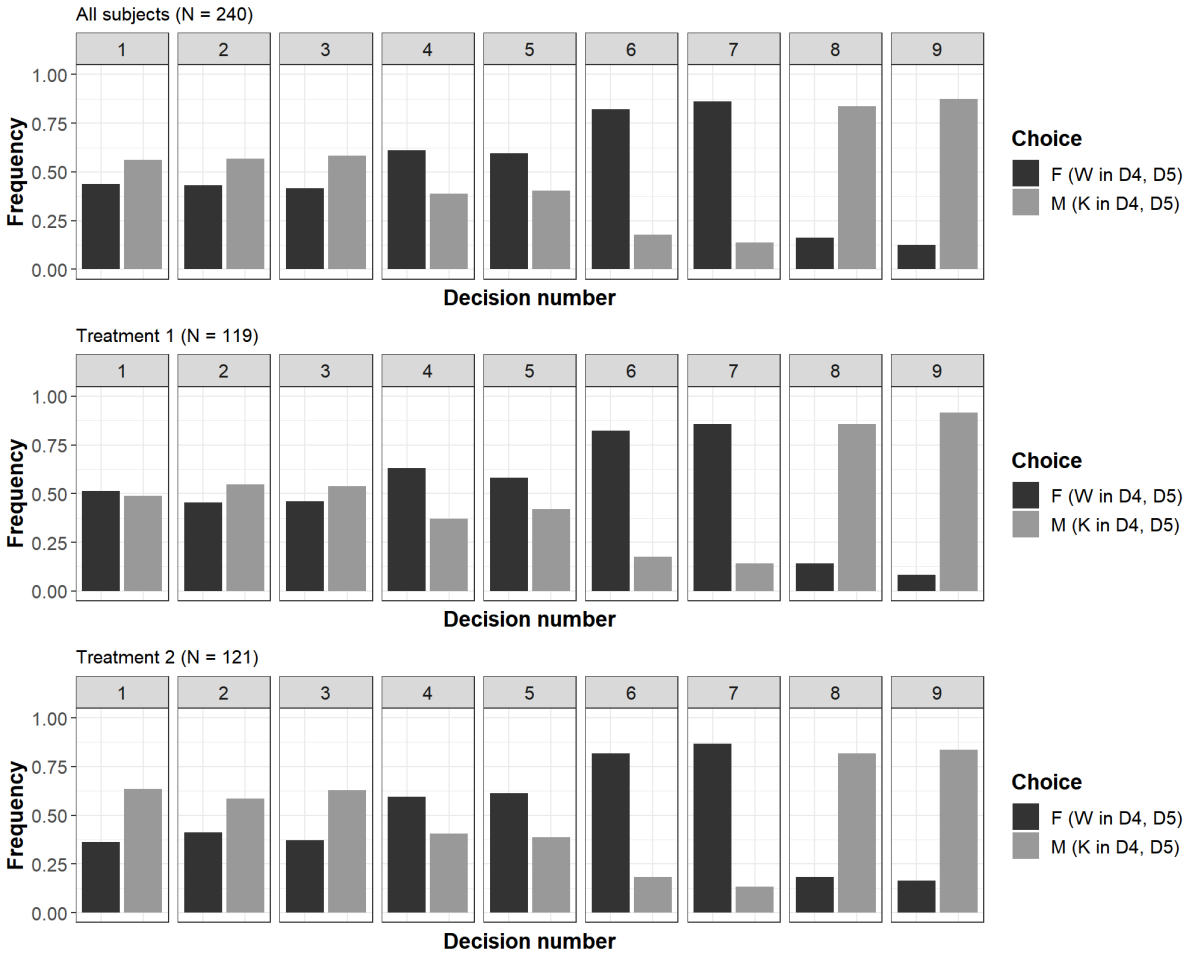


Figure 10: Beliefs of job candidates (N = 80) about gender differences in performance in the job task (the logic task). Dashed line indicates sample mean.

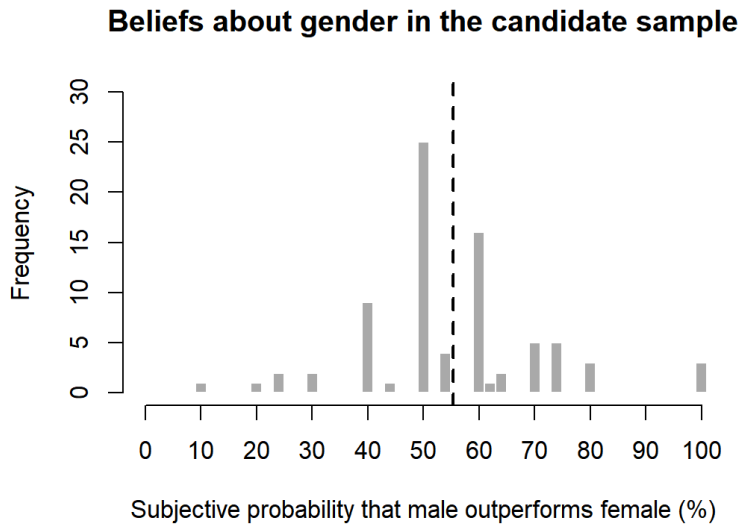


Table 6: Willingness to pay for qualifications.

	Dependent variable: Willingness to pay (Euro cents)		
	MSB excluded		MSB imputed
	(1)	(2)	(3)
T2(MW)	1.2	−2.4	−3.6
	(2.7)	(3.8)	(3.6)
Word certificate	5.5**	1.7	0.6
	(2.5)	(3.9)	(3.6)
T2(MW) * Word certificate		7.3	8.0*
		(4.9)	(4.6)
Constant	43.5***	45.4***	46.8***
	(2.4)	(3.0)	(2.8)
Observations	438	438	480
R^2	0.01	0.01	0.01
Residual Std. Error	27.2 (df = 435)	27.2 (df = 434)	26.2 (df = 476)
F Statistic	2.3 (df = 2; 435)	2.2* (df = 3; 434)	2.2* (df = 3; 476)

Notes: (i) MSB refers to responses with multiple switching behavior (MSB), (ii) Table reports OLS regressions with the results from the multiple price lists. After having taken decision 1, employers were incentivized to indicate their willingness to pay for a candidate with a certain certificate, (iii) All responses, two for each subject, are pooled, (iv) Standard errors are clustered at the employer level. *p<0.1; **p<0.05; ***p<0.01

C Lower Bound on Implicit Discrimination

Here, we explicitly derive the lower bound for implicit discrimination against female candidates. The bounds for implicit discrimination against male candidates, candidates with a knowledge certificate and candidates with a word certificate follow analogously. Recall that we denote by E the sub-population of employers choosing the female candidate in both of decisions 2 and 3.

Denote by $n_{t,E}(c_1,c_2)$ the number of employers in E in treatment $t \in \{1,2\}$ that prefer the candidate c_1 in choice set $C_1 = \{FW, MK\}$ and the candidate c_2 in choice set $C_2 = \{FK, MW\}$. Let then, for any $(i,j) \in \{1,2\}^2$ with $i \neq j$, $n^{(j)}_{t,E}(c_i) \equiv \sum_{c_j \in C_j} n_{t,E}(c_1,c_2)$ and $N_{t,E} \equiv \sum_{(c_1,c_2)\in C_1\times C_2} n_{t,E}(c_1,c_2)$, and define the corresponding fractions $\sigma_{t,E}(c_1,c_2) \equiv n_{t,E}(c_1,c_2)/N_{t,E}$ and $\sigma^{(j)}_{t,E}(c_i) \equiv n^{(j)}_{t,E}(c_i)/N_{t,E}$ in each treatment (always conditional on being in E). Now observe that, for either treatment t, the following is true:

$$\begin{aligned}
\sigma_{t,E}(MK,MW) &= \sigma^{(2)}_{t,E}(MK) - \sigma_{t,E}(MK,FK) \\
&\geq \sigma^{(2)}_{t,E}(MK) - \min\left\{\sigma^{(2)}_{t,E}(MK), \sigma^{(1)}_{t,E}(FK)\right\} \\
&= \max\left\{0, \sigma^{(2)}_{t,E}(MK) - \sigma^{(1)}_{t,E}(FK)\right\} \\
&= \max\left\{0, \sigma^{(1)}_{t,E}(MW) - \sigma^{(2)}_{t,E}(FW)\right\}.
\end{aligned}$$

Denote the lower bound thus obtained by $l^M_{t,E}$. In treatment $t = 1$, we directly observe only $\sigma^{(2)}_{t,E}(MK)$ and $\sigma^{(2)}_{t,E}(FW) = 1 - \sigma^{(2)}_{t,E}(MK)$, whereas in treatment $t = 2$, we directly observe only $\sigma^{(1)}_{t,E}(MW)$ and $\sigma^{(1)}_{t,E}(FK) = 1 - \sigma^{(1)}_{t,E}(MW)$.

In view of the random treatment assignment, assume now that, for any $(i,j) \in \{1,2\}^2$ with $i \neq j$,

$$\sigma^{(j)}_{1,E}(c_i) = \sigma^{(j)}_{2,E}(c_i).$$

This allows to identify the treatments' (then common) lower bound as

$$l^M_{1,E} = \max\left\{0, \sigma^{(2)}_{1,E}(MK) - \sigma^{(1)}_{2,E}(FK)\right\} = l^M_{2,E} \equiv l^M_E,$$

which bounds also the fraction of implicit discriminators in the entire sub-population E, $\sigma_E(MK,MW)$, since

$$\begin{aligned}
\sigma_E(MK,MW) &\equiv \frac{n_{1,E}(MK,MW) + n_{2,E}(MK,MW)}{N_{1,E} + N_{2,E}} \\
&= \frac{\sigma_{1,E}(MK,MW)N_{1,E} + \sigma_{2,E}(MK,MW)N_{2,E}}{N_{1,E} + N_{2,E}} \\
&\geq l^M_E \frac{N_{1,E} + N_{2,E}}{N_{1,E} + N_{2,E}} = l^M_E.
\end{aligned}$$

In the main text, we use simplified notation $\sigma_{1,E}^{MK} \equiv \sigma_{1,E}^{(2)}(MK)$ and $\sigma_{2,E}^{FK} \equiv \sigma_{2,E}^{(1)}(FK)$. (Analogously, for $\sigma_{1,E}^{FW}$ and $\sigma_{2,E}^{MW}$.)

Discussion Papers of the Research Area Markets and Choice 2020

Research Unit: **Market Behavior**

Research Unit: **Economics of Change**

Research Unit: Ethics and Behavioral Economics

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice