WORKING PAPER

DIVING IN THE MINDS OF RECRUITERS: WHAT TRIGGERS GENDER STEREOTYPES IN HIRING?

Hannah Van Borm
Stijn Baert

April 2022
2022/1043

GHENT UNIVERSITY

Department of Economics

Faculty of Economics and Business Administration
Campus Tweekerken, St.-Pietersplein 5, 9000 Ghent - BELGIUM

Diving in the minds of recruiters: What triggers gender stereotypes in hiring?

By Hannah Van Borm[i] and Stijn Baert[ii]

Abstract

We investigate the drivers of gender differentials in hiring chances. More concretely, we test (i) whether recruiters perceive job applicants in gender stereotypical terms when making hiring decisions and (ii) whether the activation of these gender stereotypes in recruiters' minds varies by the salience of gender in a particular hiring context and the gender prototypicality of a job applicant, as hypothesised in Ridgeway and Kricheli-Katz (2013). To this end, we conduct an innovative vignette experiment in the United States with 290 genuine recruiters who evaluate fictitious job applicants regarding their hireability and 21 statements related to specific gender stereotypes. Moreover, we experimentally manipulate both the gender prototypicality of a job applicant and the salience of gender in the hiring context. We find that employers perceive women in gender stereotypical terms when making hiring decisions. In particular, women are perceived to be more social and supportive than men, but also as less assertive and physically strong. Furthermore, our results indicate that the gender prototypicality of job applicants moderates these perceptions: the less prototypical group of African American women, who are assumed to be less prototypical, are perceived in less stereotypical terms than white women, while some stereotypes are more outspoken when female résumés reveal family responsibilities.

Keywords: hiring, gender discrimination, stereotypes, race, motherhood.

JEL-classification: J71, J16, J15, J13, J24.

---

[*] Acknowledgements. We thank Eva Derous, William Scott, and Katrin Auspurg for their insightful comments and suggestions which have helped to improve the designing of the study considerably. Next to this, we thank Didier Fouarge, Eva Van Belle, Greet Van Hoye, Bart Defloor, and my colleagues Philippe Sterkens, Eline Moens, Simon Amez, Brecht Neyt, Louis Lippens, and Joey Wuyts for the illuminating discussions and their indispensable input, which have lifted the study to a higher level. Lastly, I want to thank Eva De Vrij for her impeccable research assistance and support.

[i] Corresponding author. Ghent University. Sint-Pietersplein 6, 9000 Ghent, Belgium. Hannah.VanBorm@UGent.be. +32476807480.

[ii] Ghent University, University of Antwerp, Université catholique de Louvain, IZA, GLO, and IMISCOE.

1. Introduction

Ever since women increasingly started to enter the labour force, researchers have been concerned with studying discrepancies in employment chances between men and women. Where scholars first used regression-based methods on survey data to investigate whether women are discriminated against in hiring, in more recent decades, they have switched to experimental research methods (Guryan & Charles, 2013; Keuschnigg & Wolbring, 2016). More concretely, researchers have conducted both correspondence tests and factorial survey experiments to investigate gender discrimination in hiring. A correspondence test is a type of field experiment in which researchers send out fictitious job applications to real employers and measure positive call-back–and is by many viewed as the golden standard to measure hiring discrimination. The evidence provided by these correspondence studies is, however, eminently mixed (for overviews of these studies and corresponding results, see Baert (2018a), Kübler, Schmid, and Stüber (2018), and Lippens, Vermeiren, and Baert (2021)). While some studies find that, overall, there is no discrimination towards women (e.g., Albert, Escot, & Fernández-Cornejo, 2011; Asali, Pignatti, & Skhirtladze, 2018; Baert, De Pauw, & Deschacht, 2016; Capéau, Eeman, Groenez, & Lamberts, 2012; Carlsson & Eriksson, 2019) others find a hiring premium for women (e.g., Benhabib & Adair, 2017; Berson, 2012; Booth & Leigh, 2010; Mavlikeeva & Asanov, 2020; Valfort, 2020) and still others find a hiring penalty towards women (e.g., Busetta, Campolo, & Panarello, 2018; Galarza & Yamada, 2019; González, Cortina, Rodríguez, 2019). Moreover, the results seem to be highly heterogenous by both job applicant characteristics–e.g., very feminine versus more masculine women (e.g., Drydakis, Sidiropoulou, Bozani, Selmanovic, & Patnaik, 2018), mothers versus childless women (e.g., Brandén, Bygren, & Gähler, 2018; Bygren, Erlandsson, & Gähler, 2017; Capéau, Eeman, Groenez, & Lamberts, 2012; Correll, Benard, & Paik, 2007; Duguet & Petit, 2005; Firth, 1982; González, Cortina, & Rodríguez, 2019; Petit, 2007), or women with or without an ethnic background (e.g., Berson, 2012; Capéau, Eeman, Groenez, & Lamberts, 2012; Darolia, Koedel, Martorell, Wilson, & Perez-Arce, 2016; Duguet, Du Parquet, L'Horty, & Petit, 2015; Jacquement & Yannelis, 2012; Weichselbaumer, 2020)–recruiter characteristics–e.g., male versus female recruiters (e.g., Mavlikeeva & Asanov, 2020)–or context characteristics–e.g., male or female gender-type of job or sector (e.g., Booth & Leigh, 2010; Carlsson, 2011; Levinson, 1975; Riach & Rich, 2006; Weichselbaumer, 2004). Even between studies that focus on similar applicant-, recruiter-, and context characteristics moderation effects still seem to differ considerably in terms of size and direction.

Factorial survey experiments, on the other hand, are lab experiments in which researchers ask participants to judge short descriptions of fictitious situations, persons, or objects. In research on gender discrimination in hiring, researchers typically ask participants to rate résumés of fictitious job applicants–varying in their gender and several additional characteristics–for a hypothetical job vacancy. The results from these lab experiments are, in line with the results of the correspondence tests, very mixed (for an overview of these studies and corresponding results, see Davison and Burke, 2000; Koch, D'Mello, and Sackett (2015), and Tosi and Einbender (1985)). Where some lab experiments find, overall, no difference in the hiring chances of men and women (e.g., Bosak & Sczesny, 2011; Cunningham, Satore, & McCullough, 2010; Foschi & Valenzuela, 2008; Uhlmann & Cohen, 2007), others find that men have higher hiring chances than women (e.g., Dipboye, Arvey, & Terpstra, 1977; Shahani-Denning, Andreoli, Snyder, Tevet, & Fox, 2011). Moreover, mixed levels of heterogeneity have been found in women's hiring chances by different (i) candidate-related characteristics (e.g., their qualifications, physical attractiveness, or self-presentation; Bosak & Sczesny, 2011; Dipboye, Arvey, & Terpstra, 1977; Foschi & Valenzuela, 2008), (ii) recruiter-related characteristics (e.g., their gender; Bosak & Sczesny, 2011; Dipboye, Arvey, & Terpstra, 1977; Foschi & Valenzuela, 2008; Uhlmann & Cohen, 2007), (iii) job-related characteristics (e.g., gender type of a job; Davison & Burke, 2000), and (iv) applicant pool-related characteristics (e.g., the gender composition of the applicant pool; Heilman, 1980; Van Ommeren, Russo, De Vries, & Van Ommeren, 2005). In other words, also in vignette experimental research the moderation effects of these different applicant-, recruiter-, and context characteristics on women's hiring chances are often very different and inconsistent between studies that investigate the same or similar moderators.

In the field of economics, two prime theoretical models have been put forward in the economic literature to explain why women might be discriminated against in the hiring context: Beckers' (1957) model of taste-based discrimination and Arrow (1973) and Phelps (1972) model of statistical discrimination.[1] On the one hand, Becker (1957) argues that hiring discrimination towards women is driven by employers' (co-workers') ((customers')) negative attitudes towards them, which translate into a willingness of employers (co-workers) ((customers)) to pay a cost to avoid contact with women. On the other hand, Arrow (1973) and Phelps' (1972) model of statistical discrimination implies that (negative) gender stereotypes existing about women's abilities and skills drive the hiring discrimination against them. The reasoning behind this idea is that employers typically dispose of limited information about a job applicant when making hiring decisions (i.e., only the information mentioned on a résumé or cover letter). As a result, they will use the available information on the résumé–i.e., both job-related

[1] Note that various other theoretical models have been put forward in different academic disciplines (e.g., sociology or psychology) explaining discrimination in hiring. Although these models have each their individual take on gender discrimination in hiring, they can all be boiled down to the two main general ideas retraceable to the models of taste-based and statistical discrimination–i.e., that hiring discrimination towards women can be explained by (i) employers' tastes towards (or preferences for) a particular gender group (e.g., relational demography theory (Carolina, 2005), social categorisation and social identity theory (Tajfel & Turner, 1985)) or (ii) by the existence of gender stereotypes in society (e.g., lack of fit model (Heilman, 1983, 2001), role congruity theory (Eagly & Karau, 2002), or status incongruity hypothesis (Rudman, Moss-Racusin, Phelan, & Nauts, 2012).

(e.g., work experience and education) and nonjob-related information (e.g., gender, age, or race)–as signals of unobserved factors (e.g., job commitment and motivation). Depending on the content of the gender stereotypes, this could signal a lower level of productivity which could lead employers to refuse hiring women.

Although these two models provide us with valuable insights into why women might have lower hiring chances compared with men, they do not explain why women's hiring chances differ so much from situation to situation. A potential explanation for why these models are unable to explain the mixed results discussed above is that both models either treat employers' (co-workers') ((customers')) tastes towards women or employers' stereotypical beliefs about women as constant across social contexts. However, it could be the case that these mechanisms of hiring discrimination vary by different contextual factors (Bertogg, Imdorf, Hyggen, Parsanoglou, & Stoilova, 2020; Keuschnigg & Wolbring, 2016). To meet this limitation, researchers from outside the field of economics have developed alternative theories to explain women's varying hiring chances. One of these theories is the account developed by Ridgeway and Kricheli-Katz (2013)–which, in line with the model of statistical discrimination, argues that people's judgements of one another are driven by internalised stereotypes.

In their framework, the authors depart from the idea that in social situations–more generally–people automatically and nearly immediately categorise the other in terms of gender and that this categorisation implicitly activates various gender stereotypes in people's minds. As stated by the authors, these triggered stereotypes might, subsequently, influence people's judgements and behaviour towards one another (Macrae & Quadflieg, 2010). The degree to which gender stereotypes are triggered and bias someone's judgements depends, following Ridgeway and Kricheli-Katz (2013), on features of the context in which social interaction takes place. The first feature they mention is the salience of gender in a particular context. They distinguish two contexts in which gender is especially salient: (i) contexts that are culturally linked to a certain gender (e.g., the labour market versus the household), and (ii) contexts in which people differ in terms of gender (e.g., in a group that mainly consist out of men, being a woman will be especially salient; Berger and Webster, 2006). The second feature they mention is the extent to which a particular person is more prototypical of one of the gender categories (Macrae and Quadflieg, 2010). Meaning that if a person deviates from the prototypical image of men (i.e., the Breadwinner subtype: young white straight full-time employed fathers who are the breadwinner of the family; Edwards, 1992) or women (i.e., the Housewife subtype: young white straight mothers who stay at home to take care of their children; Rosette, Ponce de Leon, Koval, & Harrison, 2018), for example, because they are black (e.g., Goff, Thomas, & Jackson, 2008; Purdie-Vaughns, & Eibach, 2008; Sesko & Biernat, 2010) or do not have children (e.g., Rosette et al., 2018; Visser, 2002), the existing gender stereotypes related to men and women might be triggered less heavily and, as a result, may also impact people's responses towards one another differently.[2]

In the current article, we apply this broad theoretical framework to the hiring context to explain the varying hiring chances of women in the labour market. More concretely, in line with Ridgeway and Kricheli-Katz (2013)–and by that also the model of statistical discrimination–we depart from the idea that the gender stereotypes existing about women (and men) in the labour market–and more generally in society–serve as the driving forces behind employers' hiring decisions. Indeed, previous research has repeatedly mentioned that widely shared beliefs about men and women's abilities and characteristics (e.g., that men are agentic and competent and women are communal but less competent) determines women's (and men's) hiring chances (Cuddy, Glick, & Beninger, 2011; Heilman, 2012; Rosette et al., 2018).[3, 4] Therefore, we assume that also in the context of our experiment, employers will perceive male and female job applicants in stereotypical terms when making hiring decisions **(H1)**.

Moreover, in line with Ridgeway and Kricheli-Katz (2013), we argue that the degree to which these gender stereotypes are triggered depends on the circumstances under which employers make hiring decisions. First, we take into account the salience of a woman's gender in a particular hiring context. In line with Heilman (1980), we argue that salience of gender is "established by varying the distinctiveness of femaleness as characteristics" (Heilman, 1980, p.388) Consequentially, we consider the following two factors to influence the salience of a job applicant's gender in a hiring context: (i) the gender type of the job for which job candidates apply and (ii) the gender composition of the applicant pool from which employers assess job applicants. Indeed, as argued by Heilman (1983, 1995), many jobs in society are associated with a particular gender, either because men or women are disproportionally represented in these jobs or because the job responsibilities are culturally associated with a particular gender

[2] Note that the theoretical framework of Ridgeway & Kricheli-Katz (2013) is developed in the context of the United States. As a result, the prototypical images of men and women and associated gender stereotypes discussed in their framework are those of traditional American or Western men and women. However, these prototypical images and associated stereotypes are not necessarily the same in other countries.

[3] These stereotypical views about the attributes and skills of men and women in the labour market originate from the traditional gender roles of men as breadwinners and women as household caretakers deeply embedded in Western society (Rosette et al., 2018).

[4] In the context of the labour market, the concept of agency typically covers achievement-orientation (e.g., competence, ambition, and task-orientation), the tendency to take charge (e.g., assertiveness, dominance, forcefulness), autonomy (e.g., independence, self-reliance, decisiveness) and rationality (e.g., analytic skills, logic reasoning, objectiveness). In contrast, communality has come to denote being concerned for others (e.g., kindness, caring, consideration), affiliative tendencies (e.g., warmness, friendliness, collaboration), deference (e.g., obedience, respectfulness, self-efficacy) and emotional sensitivity (e.g., perceptiveness, intuition, understanding) (Heilman, 2012).

(Heilman, 2012; Lyness & Heilman, 2006).[5, 6] Due to its gendered nature, the type of job for which an applicant applies can, thus, influence the degree to which someone's gender is salient to employers, and as a result, also the extent to which gender stereotypes in the minds of these employers are triggered when making hiring decisions. Following this line of thought, we hypothesise that the stereotypical perceptions about the attributes and skills of women will be triggered more heavily in male gender-typed jobs **(H2)**. Furthermore, since–in line with Heilman (1980)–we argue salience of gender to be established when femaleness as characteristic is more distinctive, we expect that women's gender will be especially salient in applicant pools that consist primarily out of male job applicants because in these situations being female is in stark contrast to the 'male-norm'. Consequentially, we hypothesise that stereotypes about women will be triggered more heavily in male-dominated applicant pools **(H3)**.

Second, next to the salience of gender in the hiring context, we also consider the gender prototypicality of a job applicant. Given that the United States' population has become more and more diverse and the number of women in the labour force has increased remarkably over the past decades (i.e., from 20.30% in 1920 to 47.30% in 2019; U.S. Bureau of Labor Statistics, 2021b), the population of women in the workforce also became increasingly diverse. Consequentially, today, employers are confronted with a variety of women differing in terms of age, race, parental status, and many other characteristics.[7] It, therefore, seems valid to believe that also in the hiring context some women deviate from the prototypical image of women embedded in western societies (i.e., young, white, straight mothers). In our study, we focus on two subgroups of women varying in prototypicality: black women and women with family responsibilities. In line with Ridgeway and Kricheli-Katz (2013), we argue that black women deviate from the prototypical image of women due to their skin colour and, therefore, will be perceived as un-prototypical for their gender category. This claim is supported by the results of previous empirical research that found black women to be perceived as less prototypical for their gender category than white women (e.g., Merriweather, 2020; Sesko & Biernat, 2010). Following this line of thought, we hypothesise that the stereotypes about women are triggered less when the female job applicant is black **(H4)**. On the contrary, we argue that women with many family responsibilities can be perceived as more prototypical for their gender category since being a care giver coincides with the prototypical image of

[5] An example of a job or sector in which men are disproportionally represented (i.e., male-dominated jobs) are jobs in the prestigious STEM (science, technology, engineering, mathematics) fields. Moreover, an example of a job or sector in which women are overrepresented (i.e., female-dominated jobs) is secretary- or administrative work (U.S. Bureau of Labour Statistics, 2021a, 2021b)

[6] An example of an occupation of which the job responsibilities are culturally linked to men (women) are leadership positions (nursing positions) (Clow, Ricciardelli, & Bartfay, 2015; Schein, 1973, 1975).

[7] Indeed, according to data of the U.S. Bureau of Labor Statistics from the Current Population Survey, the number of women belonging to various races and ethnicities active on the US labour market has increased remarkably over the past decades, as well as the proportion of women with children and the number of women between the ages of 16 and 65 (U.S. Bureau of Labor Statistics, 2021b).

women (e.g., Visser, 2002). Consequentially, we hypothesise that the stereotypes about women will be triggered more heavily when the female job applicant explicitly mentions family responsibilities in her résumé **(H5)**.

To empirically test the hypotheses formulated above, we conduct a vignette experiment in the United States with people with genuine experience in hiring. In our vignette experiment, we manipulate, as mentioned above, four distinct factors. More concretely, we ask participants to assess multiple fictitious job applicants (i.e., five fictive candidates varying–amongst other things–in race **(F1)** and family responsibilities **(F2)**) from a particular pool of job candidates (i.e., one out of 80 applicant pools differing in gender composition **(F3)**) for a specific hypothetical job vacancy (i.e., one out of twelve possible job vacancies varying in gender type **(F4)**). The participants had to evaluate the candidates in terms of their hireability and 21 statements corresponding to different stereotypes about women in the labour market.

In doing so, we contribute to the existing literature by conducting a vignette experiment in which we survey participants' stereotypical beliefs of job applicants directly and examine how these stereotypes–activated in the minds of recruiters–might be triggered to varying extents in different hiring contexts, considering both the salience of gender in a hiring context and the degree to which a job applicant is prototypical for her gender category–as argued by Ridgeway and Kricheli-Katz (2013). By investigating the possible moderator effects of different contextual factors on stereotype activation–to the best of our knowledge for the first time–we gain deeper insights into why women's hiring chances differ so much from situation to situation.

Furthermore, we also contribute methodologically by conducting a name categorisation experiment (see Subsection 2.3) to validate the quality of job applicants' names as signals for race and gender. Therefore, meeting the valid criticism formulated by Gaddis (2017, 2019) that the names used in experimental research are often bad signals for race (or gender) since they might not signal the race (or gender) intended or might also signal other aspects such as social class.

Next to its academic relevance, our research is also relevant from a societal point of view. Knowing when and to what degree gender stereotypes affect women's hiring chances is crucial information to counter barriers experienced by women in the labour market. Indeed, if women know which stereotypes are triggered and are affecting their hiring chances in particular contexts, they can proactively take them into account when applying for a job and, as a result, may potentially increase their hiring chances by doing so.

2. The vignette experiment

As mentioned in the previous section, we conduct a vignette experiment to investigate whether the degree to which stereotypes about women are triggered and influence employer's hiring decisions differs by the dimensions mentioned in the theoretical framework of Ridgeway & Kricheli-Katz (2013) (i.e., the salience of gender in a hiring context and the gender prototypicality of a job applicant). A vignette experiment is an application of the factorial survey method, developed by Rossi (1951) and is often used to study human attitudes, judgements, and behaviour intentions (Auspurg & Hinz, 2014; Rossi & Nock, 1982). As mentioned in the previous section, in these experiments, participants typically judge short descriptions of fictitious situations, persons, or objects depicted in vignettes. The characteristics of these vignettes (i.e., vignette dimensions) vary systematically or randomly over a pre-defined number of categories (i.e., vignette levels) (Sauer, Auspurg, Hinz, & Liebig, 2011). This experimental design is then implemented in a(n) (online) survey. The main advantage of this research method is that one can avoid both the disadvantages related to lab experimental and survey research to some extent. More concretely, the often mentioned disadvantage of low external validity of lab experiments can somewhat be avoided by the implementation of an experimental design in a(n) (online) survey. Indeed, (online) surveys typically allow for reaching relatively large and heterogeneous populations (Auspurg & Hinz, 2014; Damelang & Abraham, 2016; Van Belle, Di Stasio, Caers, De Couck , & Baert, 2018; Wallander, 2009). Moreover, by experimentally manipulating the vignette levels of the different vignette dimensions, one assures internal validity and avoids the common issue of confounding variables in survey research (Auspurg & Hinz, 2014; Damelang & Abraham, 2016; Van Belle et al., 2018; Wallander, 2009).

For research on hiring discrimination, in particular, vignette experiments also have an advantage over the often conducted correspondence tests. Although correspondence tests are the golden standard to measure hiring discrimination, they are less appropriate to gain insights into the underlying mechanisms of this discrimination. Indeed, correspondence tests only allow researchers to investigate employers' binary decisions to invite a job candidate to a job interview or not. Vignette experiments, in contrast, allow researchers to survey employers regarding a multitude of decisions and the motivations behind them. Consequentially, vignette experiments are more appropriate to research the underlying mechanisms behind employers' hiring decisions–i.e., the prime goal of this research–than correspondence tests (Van Belle et al., 2020; Van Borm, Burn, & Baert, 2021).

It is not surprising, then, that vignette experiments have been conducted extensively in various academic fields, including economics (e.g., Ambuehl & Ockenfels, 2017; Eriksson & Kristensen, 2014;

Mathew, 2017), sociology (e.g., Auspurg, Hinz, & Sauer, 2017; Jasso, 2006; Liebe, Moumouni, Bigler, Ingabire, & Bieri, 2020; McDonald, 2019; Rivera & Tilcsik, 2016; Wallander, 2009), and psychology (e.g., Derous, Nguyen, & Ryan, 2009; Derous, Ryan, & Nguyen, 2012; Webster, O'Toole, O'Toole, & Lucal, 2005). Moreover, they have been widely used to study employers' hiring decisions, as well as potential hiring discrimination towards various minority groups in the labour market (e.g., Baert & De Pauw, 2014; Damelang, Abraham, Ebensperger, & Stumpf, 2019; Derous et al., 2009; Derous et al., 2012; Di Stasio, 2014; Hosoda, Stone, & Stone-Romero, 2003; McDonald, 2019; Van Belle et al., 2018; Van Belle, Caers, De Couck, Di Stasio, & Baert, 2019; Van Belle et al., 2020; Van Borm, Dhoop, Van Acker, & Baert, 2020; Van Hoye & Lievens, 2003).

In what follows, we first discuss the experimental design used in our study (Subsection 2.1) and subsequently describe the online survey in which the experimental design is implemented (Subsection 2.2).

2.1. Vignette design

In the current study, the vignettes implemented in the online survey consisted of short descriptions of fictitious job candidates differing in five distinct applicant characteristics (i.e., the vignette factors). More concretely, we presented the name of a job applicant, his or her commuting distance, his or her experience in the occupation, a potential employment gap and the reason for this gap, and his or her extracurricular activities. We presented this information to participants in small tables.[8,9] The five applicant characteristics varied over four to twelve levels. Moreover, we selected all applicant characteristics and corresponding levels such that no illogical or implausible cases could occur. We present an overview of the five vignette dimensions and the corresponding levels in Table 1.

<Table 1 about here>

[8] The choice to vary the candidate characteristics over five vignette factors was made based on the recommendation of Auspurg and Hinz (2014) to work with vignettes of midlevel complexity, i.e., vignettes in which approximately seven (plus or minus two) vignette dimensions are varied. By using a midlevel number of vignette dimensions, we avoid participants to become overburdened by a too complex vignette design and, at the same time, assure participants to be stimulated enough not to drop out because of boredom or fatigue, which might happen with an overly simple design in which participants have to rate several very similar vignettes.

[9] As recommended by Auspurg and Hinz (2014), we work with tabular vignettes instead of text vignettes because these are better suited to decision tasks involving lists of decision criteria, such as evaluating fictitious résumés. Indeed, tabular vignettes' more straightforward presentations of vignette factors help participants to form more consistent judgements, which is especially useful in the context of our experiment in which participants have to evaluate multiple fictitious job applicants. Moreover, previous research has shown that tabular vignette designs produce similar evaluations to text designs (Auspurg & Hinz, 2014; Shamon, Dülmer, & Giza, 2022; Sauer, Auspurg, & Hinz, 2020).

As mentioned in Section 1, one of the dimensions we vary in our experiment is the gender prototypicality of a job candidate. Based on the results found in previous research on gender prototypicality (e.g., Goff, Thomas, & Jackson, 2008; Purdie-Vaughns, & Eibach, 2008; Sesko & Biernat, 2010; Visser, 2002), we decided to manipulate the prototypicality of the female job candidates in two ways: (i) through job applicants' race **(F1)** and (ii) family responsibilities **(F2)**. We signalled job applicants' race (i.e., black or white)–as well as their gender (i.e., male or female)–through their names. As becomes clear from Table 1, we decided to work with 12 different name levels (i.e., four corresponding to white men, four to white women, two to black men and two to black women, resulting in four distinct gender-race groups). In doing so, the composition of our applicant pool reflected the one of the United States population. Indeed, about 60% of people in the United States is 'only white', and about 50% is female (U.S. Census Bureau, 2021b). To avoid potential name effects, we selected multiple names (i.e., four names for both black men and women and eight names for both white men and women) for each of the four gender-race groups, resulting in 24 distinct names in total. When filling in the survey, participants got randomly assigned one of them from the pool of possible names corresponding to their job applicants' gender-race group. To ensure the names only signalled the intended genders and races (and not, for example, differences in social class), all 24 'black' or 'white' names were carefully selected and pretested in a name categorisation study before implementing them in the current experiment. We will discuss the selection process and testing of the different names in Subsection 2.3, where we describe the name categorisation study in more detail. We present an overview of the, in total, 24 combinations of first and last names used in the experiment in Table A–1 in the Appendix.

Besides, we signal women's caretaking responsibilities to employers through the employment gap mentioned in the vignettes (i.e., employment gap due to family responsibilities versus no employment gap, an employment gap due to unemployment, or an employment gap due to health issues).[10] We decided to signal one's caretaking responsibilities in this way because of its realism. Indeed, in today's labour market the majority of individuals experience a lapse in employment at some point in their careers, most often due to unemployment, care for children or other family members, or health issues (Weisshaar, 2018). Moreover, we selected the levels of this vignette dimension based on previous research (e.g., Bright & Davies, 1999; Eriksson & Rooth (2014), Kleist, 2001; Namingit, Blankenau, & Schwab, 2021; Weisshaar, 2018).

As described in Table 1, we also experimentally manipulate (i) the distance someone needs to commute between their home and the presented job, (ii) one's experience in the occupation, (iii), and someone's

[10] We clarified the duration and period of the employment gap in the scenario sketch offered to the participants in the online survey (see Subsection 2.2).

extracurricular activities. We included these additional vignette dimensions to mimic real-life hiring decisions as closely as possible and cover up the prime goal of the research to avoid social desirability bias. These extra vignette dimensions are all elements typically mentioned on U.S. résumés and were selected based on the findings of previous studies (i.e., Olian, Schwab, & Haberfeld, 1988; Lahey, 2008; Nuijten, Poell, & Alfes, 2017; Carlsson, Reshid, & Rooth, 2018; Van Belle et al., 2018; Van Belle et al., 2019; Van Borm, Burn, & Baert, 2021). Moreover, to check whether these extra factors were perceived to be (i) relevant, (ii) realistic, and (iii) informative for employers, we conducted a pilot test of our survey with 80 U.S. Prolific-users experienced in hiring.[11] We will come back to this pilot test in Subsection 2.2.

Combining all vignette levels for the five vignette dimensions (i.e., 12 × 4 × 4 × 4 x 4) resulted in 3,072 unique vignettes (i.e., the vignette universe). Following Auspurg & Hinz (2014), we decided to sample a number of vignettes from this universe (i.e., a vignette fraction) using a D-efficient Resolution V design.[12] A D-efficient Resolution V design selects vignettes with the most statistical power while ensuring the estimation of all relevant parameters (Auspurg & Hinz, 2014).[13, 14, 15] In total, we sampled 400 vignettes from the vignette universe using Kuhfelds' (2010) computer algorithm, resulting in a D-efficient Resolution V design with a sufficiently high D-efficiency of 92.854, which, at the same time, allows for the estimation of all main- and two-way interaction effects.[16]

Moreover, in line with Ridgeway and Kricheli-Kayz (2013), we also consider the salience of gender in a particular hiring context as a potential explanation for the varying hiring chances of women in the labour market (see Section 1). As discussed in Section 1, one way through which we manipulate the salience of gender in our experiment is through the gender composition of the applicant pool **(F3)** from which

---

[11] Prolific Academic is an online crowdsourcing platform on which scholars can recruit subjects to perform particular tasks in return for financial compensation (Palan & Schitter, 2018).

[12] We decided to sample a vignette fraction from the vignette universe because letting participants evaluate 3,072 distinct vignettes sufficiently often (i.e., at least five times, Auspurg & Hinz, 2014) would require a major sample of participants or having each participant assess an enormous number of vignettes which could cause fatigue among them and compromise the quality of our data (Auspurg & Hinz, 2014). By sampling a vignette fraction from the universe, we avoided these issues.

[13] Sampling a vignette fraction from the vignette universe typically leads to a loss of information and could cause partial confoundings to occur (i.e., the occurrence of moderate or strong correlations between different vignette dimensions; Atzmüller & Steiner, 2010; Auspurg & Hinz, 2014). However, using a D-efficient Resolution V design–instead of random sampling–we manage to retain much of the available information and avoid partial confoundings.

[14] A D-efficient design enhances statistical precision by maximising both orthogonality and level balance (i.e. equal frequencies of all levels).

[15] The resolution of a design defines which effects, including possible interaction effects, are identifiable with an employed vignette fraction (Dülmer, 2007; Kuhfeld, 1997; Auspurg & Hinz, 2014). Resolution V corresponds to a design where all main effects and two-way interactions are identifiable.

[16] To select the 400 vignettes, we used the freeware macro %Mktex developed by Kuhfeld (2010). Considering the number of vignettes one wants to use in the experiment, the parameters one tends to identify, and the number of factors and associated levels one defines, this algorithm first builds a set of potential designs and, subsequently, searches for the design with the highest D-efficiency. D-efficiency is sufficiently high when it exceeds 0.90 (Auspurg &Hinz, 2014; Kuhfeld, 1997; Kuhfeld, Tobias, & Garratt, 1994). For more information, we refer to Auspurg and Hinz (2014).

employers evaluate job applicants. To operationalise the gender composition of an applicant pool in our experiment, we created 80 different applicant pools varying in the number of men and women included. We constructed the different applicant pools by blocking the 400 unique vignettes discussed above in 80 decks of five job applicants each, again using Kuhfeld's (2010) computer algorithm as discussed by Auspurg and Hinz (2014).[17] The number of women in an applicant pool varied between one and four.[18] During the experiment, we randomly assigned one of the 80 decks to each of the participants.

2.2. Online survey

As mentioned at the beginning of Section 2, one of the prime features of factorial survey applications is the incorporation of a multidimensional experimental design within a(n) (online) survey (Auspurg & Hinz, 2014). In our study, we incorporated the experimental design described in the previous subsection in an online survey designed in Qualtrics and administered in English.

The survey consisted of two parts: (i) a non-experimental survey and (ii) the vignette experiment. In the non-experimental survey, we questioned participants regarding twelve different personal characteristics.[19, 20] More concretely, we surveyed them about five demographic characteristics: (i) their gender (male, female, or 'X'), (ii) their age, (iii) the state they live in, (iv) their highest obtained degree (university education, higher education outside the university, secondary education, or lower than secondary education), and (v) their race-ethnicity (white, Hispanic or Latino, black or African American, Asian, American Indian or Alaska Native, Middle Eastern or North African, Native Hawaiian or other Pacific islanders, or some other race or ethnicity).[21]

Moreover, we surveyed participants' tendency to respond in a socially desirable manner via the 13-item version of the Marlowe-Crowne Social Desirability Scale (MC-SDS) developed by Reynolds (1982)–one of the instruments used most to measure social desirability (Beretvas, Meyers, & Leite, 2002; Sârbescu,

[17] We decided to limit the size of the applicant pools to five unique job applicants per participant for two reasons. First, we considered the recommendation of Auspurg and Hinz (2014) to use not more than ten vignettes per participant. Second, we decided to work with only five vignettes per participant because we wanted to avoid an excessive survey time, learning effects, and fatigue among the participants.

[18] To avoid order effects, we randomised the order of the five different job applicants within an applicant pool.

[19] Before we redirected the participants to the actual survey, we showed them a short introduction in which we explained the aim of the research and described the task they would have to perform. We described the aim of the study broadly and vaguely to make sure the actual goal of the study was not too clear to participants and avoid socially desirable answering. Moreover, we mentioned the estimated survey time, ensured anonymity and confidentiality of their answers, and highlighted the value of their participation. After the introductory text, participants had to read and sign a consent form to participate in the survey.

[20] We included these questions in our online survey to be able to conduct different robustness analyses.

[21] Next to indicating their highest obtained degree, the participants also had to check an additional box by way of an attention check.

Costea, & Rusu, 2011; Baert, 2018b). The scale consists of 13 statements describing culturally sanctioned or approved behaviour (e.g., 'There have been occasions when I took advantage of someone.'). Participants had to indicate whether the statements applied to them or not using 'true' or 'false' checkboxes. We coded the answers so that socially undesirable responses coincided with a score of 0 and socially desirable ones corresponded to the number 1. Cronbach's alpha for this scale is 0.834 for our sample–which is sufficiently high. Summing the scores for all items resulted in a total score for socially desirable answering between 0 and 13. In line with Van Borm, Burn, & Baert (2021), we divided this number by 13 to obtain a proportion between 0 and 1.

Lastly, we also asked participants to answer six questions related to their experience with evaluating job applicants. More concretely, they had to indicate (i) whether they had experience in evaluating job candidates in their current or last job (yes or no), (ii) the frequency with which they were, in general, involved in the evaluation of job candidates in their current or last job (daily, weekly, biweekly, monthly, once per semester, once a year, or less frequently), (iii) the frequency with which they were involved in the evaluation of job candidates in the last year (0 times, 1 time, 2 times, 3 times, 4 times, or 5 times or more), (iv) the amount of time (in years) they had been involved in evaluating job candidates (less than one year, 1–5 years, more than 5 years, or not applicable), (v) whether they recruited in a specific sector (yes or no) and if so which one (construction, technical sector, ICT sector, financial sector, administration, sales, human resources, transport or logistics, or other), and (vi) a description of their current or last occupation (manager, specialist in personnel and career development, employment agency employee, management assistant, general administrative assistant, or other).

The second part of the survey comprised of the vignette experiment discussed in Subsection 2.1. At the beginning of the online experiment, we provided participants with some experimental instructions. More concretely, we asked participants to imagine themselves to be a recruiter for a particular (hypothetical) firm for which they had to fill in a certain vacancy.[22] Subsequently, we showed them the description of the vacancy. As mentioned in Section 1, next to manipulating the gender composition of the applicant pool, we also manipulated the gender type (i.e., male or female gender-typed) of the different job vacancies **(F4)** to vary the salience of gender in a particular hiring context in our experiment. More concretely, we constructed twelve different job vacancies–i.e., the job of (i) ophthalmic laboratory technician, (ii) welder, (iii) payroll and timekeeping clerk, (iv) production, planning, and expediting clerk, (v) door-to-door sales worker, (vi) food-batch-maker, (vii) operations research analyst, (viii) chemical engineer, (ix) proof-reader, (x) human resource manager, (xi) insurance

[22] We decided to let the participants evaluate job candidates for a hypothetical firm instead of their own firm to ensure the internal validity of our experiment.

sales agent, and (xii) wind farm support specialist–that differed in terms of (i) the representation of men and women in a job, and (ii) the extent to which the job requirements are typically linked to a particular gender. More precisely, we varied (i) the proportion of men or women typically employed in the job, (ii) the required skill level (iii) the required leadership skills necessary to perform the job, (iv) the level of physical effort, and (v) the level of customer contact associated with the job. Based on a thorough reading of the relevant literature, we assume that (i) jobs in which men are overrepresented, (ii) high skilled jobs, (iii) jobs that require much leadership skills, and (iv) physically demanding jobs will be perceived as jobs with a male-gender type, while (i) jobs in which women are overrepresented and (ii) are associated with high levels of customer contact will be perceived to be female gender-typed jobs (e.g., Galinsky, Hall, and Cuddy, 2013; Heilman, 2012, Moshavi, 2004; Rosette et al., 2018). To increase the external validity of our experiment, we selected jobs from different sectors. We present an overview of the different jobs and corresponding job characteristics in Table A–2 in the Appendix.

We selected the jobs based on data found on O*Net, as well as data from the American Community Survey provided by IPUMS USA (Ruggles et al., 2021).[23] Moreover, we created the descriptions of the different job vacancies based on the descriptions found on O*Net. To avoid any effects of the descriptions on our results, we formulated them as uniformly as possible. An overview of the different descriptions can be found in Table A–3 in the Appendix. We randomly assigned the job vacancies to the participants in such a manner that all twelve vacancies were presented with equal probability (and did not correlate with the deck of fictitious profiles assigned).

After reading the job description, participants had to fill in a comprehension check. We included this check to examine whether the participants' perceptions about their assigned job matched the objective job characteristics found on O*NET and IPUMS USA. Moreover, we surveyed participants' experience with evaluating job applicants for the presented job vacancy and their feeling of competence to perform the task at hand.

Once participants completed the comprehension check, they were redirected to a scenario sketch. In the scenario sketch, we provided the participants with some additional instructions. More concretely, they were told that they had to give a second opinion regarding five job candidates with which a colleague had had a first face-to-face interview. They had to formulate their advice based on their colleague's interview notes (summarised in small tables presented on the following pages). Moreover, we mentioned that all applicants were formally eligible for the job in terms of education and work experience and clarified the duration (i.e., between one and six months) and period (i.e., over the past

[23] O*Net is an online databank developed by the U.S. Department of Labor/Employment and Training Administration summarising occupational information on thousands of jobs (National Center for O*NET Development, 2019).

five years) of the employment gaps mentioned in the tables. Furthermore, we informed the participants that they had to rate each candidate accurately and based on their own insights.

Next, we showed the participants the tabulated summaries of the five job applicants.[24] Based on the information given, they first had to indicate the probability with which they would invite the job candidate to the next round (i.e., hereafter 'invitation probability') using an 11-point Likert scale going from 0 ('strongly disagree') to 10 ('strongly agree'). We adopted the statement from Sterkens, Baert, Rooman, and Derous (2021). Next, participants had to rate the applicants concerning 18 statements corresponding to different stereotypes existing about women in society, as well as three additional statements that surveyed participants willingness to collaborate with the job applicant and their perceptions about the willingness of co-workers and customers to interact with the presented job candidates. We created the 18 statements based on a thorough reading of the relevant literature and formulated them in analogy with Van Borm, Burn, and Baert (2021). Likewise, the three additional statements surveying participants willingness to collaborate with the job applicants were adopted from Van Borm, Burn, and Baert (2021). An overview of the different statements and associated stereotypes and attitudes can be found in Table 2.

<Table 2 about here>

As mentioned in Section 1, we hypothesise that women will be more associated with communal attributes and soft skills (e.g., Cuddy, Fiske, & Glick, 2004; Fiske, Cuddy, Glick & Xu, 2002) and less with agentic characteristics or hard skills (e.g., Cuddy, Fiske, & Glick, 2004; Fiske, Cuddy, Glick & Xu, 2002). More concretely, based on a thorough reading of the relevant literature, we expect women to be perceived as having (i) more social abilities and to be more (ii) supportive, (iii) more collaborative, (iv) more open towards new people and experience, (v) and more creative (e.g., Heilman, 2012; Niemann, Jennings, Rozelle, Baxter, & Sullivan, 1994). In contrast, we assume them to be perceived as being less (vi) assertive, (vii) less autonomous, (viii) less ambitious, (ix) less resilient, (x) less , and having (xi) more respect towards authority, (xii) a lower sense of responsibility, (xiii) less intellectual abilities, and (xiv) less physical abilities (e.g., Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Collins, 2000, 2004; Cuddy, Fiske, & Glick, 2004; Donovan, 2011; Fiske, Cuddy, Glick & Xu, 2002; Galinsky, Hall, & Cuddy, 2013; Ghavami & Peplau, 2013; Heilman, 2012; Plant, Hyde, Keltner, & Devine, 2000; Rosette et al., 2018; Rudman & Glick, 1999; West, 2008; White, 1985). Furthermore, because women are often perceived as the family's caretaker, we also assume women to be perceived as (xv) less motivated, (xvi)

[24] We displayed the candidate features in the same order as they would appear in real résumés, i.e., the same order as in Table 1. Moreover, to ensure participants consider all elements of information mentioned in the tables and avoid order effects, we fixed the position of each of the tables on the screen such that it remained visible to the participants while assessing the candidates

less reliable, (xvii) less flexible, and (xviii) more absent due to family responsibilities (e.g., Morgan, Walker, Hebl, & King, 2013; Rosette et al., 2018). Lastly, because women are in general perceived to be more communal, we believe that participants, co-workers, and customers (in the minds of the participants) will be more willing to cooperate with women than men.[25]

Lastly, after evaluating the five job applicants, participants were presented with two manipulation checks. First, they had to guess what the goal of the exercise they just performed was so we could check whether the participants figured out that the research was actually about hiring discrimination towards women, which could increase their tendency to answer in a socially desirable way. They had to formulate their answer in an open-ended question. Next, to investigate whether the participants received the experimental manipulation and interpreted the gender and race signals–signalled through the names of the job applicants–correctly, they first had to indicate what the name of the different job applicants was and, subsequently, had to specify with which (i) gender, (ii) race, (iii) religion, and (iv) social class they associated that name. Besides, they also had to indicate on an 11-point Likert scale going from 0 ('strongly disagree') to 10 ('strongly agree') whether they found it hard to remember the names of the job applicants when asked.

As mentioned in Subsection 2.1, we ran a pilot study on Prolific with 80 U.S. citizens who were experienced in evaluating job applicants to examine whether participants thought the online survey to be clear and well-constructed. More concretely, the participants had to fill in the entire online survey as well as an evaluation form in which they had to rate (i) the profiles of the job applicants, (ii) the job vacancy, and (iii) the overall survey on their clarity, relevance, ecological validity, and informativity. The results of this pilot with 400 (i.e., 80 × 5) candidate evaluations, which are available upon request, indicated that both the quality of the profiles of the job applicants, the job vacancies, and the overall survey was sufficiently high. Moreover, the pilot also indicated that the names used in the experiment were good signals for gender, race, social class and religious affiliation. Indeed, in the context of our pilot test, most of the participants categorised the names in the correct categories in the manipulation check mentioned above.

In what follows, we discuss the name experiment conducted to select high-quality names as signals for someone's gender and race.

[25] Note that the results of previous research on stereotypes about women find mixed results concerning women's perceived intellectual abilities and ambition. Indeed, although some studies argue that women are perceived to be less competent (e.g., Cuddy, Fiske, & Glick, 2004; Fiske, Cuddy, Glick & Xu, 2002) and less ambitious (e.g., Heilman, 2012), other studies find the opposite (e.g., Landrine, 1985; Weitz & Gordon, 1993).

2.3. Name categorisation experiment

As mentioned in Subsection 2.1, we conducted a name testing study to guarantee the quality of the names as signals for gender and race (and consequentially also the experimental manipulation of job applicants' gender prototypicality). We conducted the name experiment with 50 U.S. Prolific users. In the experiment, participants had to categorise 20 combinations of white or black first and last names in different gender, race, social class, religious affiliation, and immigrant generation categories (i.e., 1,000 observations in total).

In doing so, we meet the justified criticism of Gaddis (2017, 2019) that the names used in many of the existing experimental studies are noisy signals for race because (i) they might not signal the race intended and (ii) they might also signal other aspects than race, such as social class. Indeed, in his research with 7,936 U.S. Amazon Turk users, Gaddis (2017, 2019) found that not all names used in previous studies are good signals for race (i.e., black versus white) and that many of the names used are associated with a particular social class (i.e., blacks are often associated with working or low class, while white names are more often associated with middle and high class).[26]

By conducting this name categorisation experiment, we, thus, improve the internal validity of our vignette experiment remarkably over other experimental research that typically does not take into account the potential noise of names as signals for race. Moreover, we also add to the existing literature by supplementing Gaddis' (2017, 2019) research. That is, we investigate not only people's gender, race, and social class associations regarding various names–as is done by Gaddis (2017, 2019)–but also potential associations between different names and various religious affiliations and immigrant generations. Investigating these two additional characteristics is relevant for two reasons. First, previous research, indeed, has found that certain names are associated with particular religious affiliations and immigrant generations. More concretely, Van Borm and Baert (2021) found that, for example, Arab names–such as Abdullah Malik or Samira Mohammed–are often associated with Islam. Moreover, they found that combinations of ethnic-sounding first and last names–such as Pheng Chan or Mei Lin–are associated more with first generation immigrants, while combinations of Anglo-sounding first names with ethnic-sounding surnames–such as George Yang or Susan Wong–are typically associated with second, third, or higher-generation immigrants.

Second, over the years, researchers also found compelling evidence that both someone's religious affiliations and immigration generation impacts one's hiring chances (e.g., Acquisti & Fong, 2020; Busetta, Campolo, & Panarello, 2018; Carlsson, 2010; Di Stasio, Lancee, Veit, & Yemane, 2021; Drydakis,

[26] Similar to Prolific Academic, Amazon Mechanical Turk is an online crowdsourcing platform via which researchers can recruit subjects to perform certain tasks in return for financial compensation (Palan & Schitter, 2018).

2010; Koopmans, Veit, & Yemane, 2019; Oreopoulos, 2011; Valfort, 2020; Veit & Thijsen, 2021; Wallace, Wright, & Hyde, 2014; Yemane, 2020). It, therefore, seems valid to believe that the names used in previous experimental research might also signal unintended religious and immigrant generation signals and, thus, bias the results found to some extent.

To select the first names for our name categorisation experiment, we used the research of Gaddis (2017, 2019) as a starting point. More concretely, from Gaddis' (2019) validated list of first names, we selected 61 of them that corresponded with the intended race and gender categories while ensuring they did not differ in terms of social class: i.e., 17 'white' male first names, 21 'white' female first names, 13 'black' male first names, and ten 'black' female first names. Moreover, to check the potential age signal of Gaddis' (2017, 2019) validated list of first names, we used the USA Social Security Administration's database to check the popularity of the 61 first names in a certain period of time (Social Security Administration, 2021). We only considered those first names that were popular between 1985-1990 to use in our name experiment. In the end, we selected 48 first names to include in the name categorisation test: 16 male and 16 female 'white' first names and eight male and eight female 'black' first names.

Furthermore, to select 48 corresponding family names, we used U.S. Census data on common surnames by ethnicity (U.S. Census Bureau, 2021a). The U.S. Census database gives an overview of the number of people with a particular last name in the United States in 2010 and ranks them by the frequency with which they occur. Moreover, it provides information on the percentage of people belonging to a particular race-ethnicity group carrying a specific last name. Only last names that occurred frequently in the 2010 U.S. Census and that were predominantly carried by one of the two race-groups were selected to be implemented in the name categorisation experiment. We randomly paired the different surnames with one of the chosen first names.

As mentioned in Subsection 2.1, we selected 24 out of the 48 full names to use in our actual vignette experiment based on the results of the name categorisation study–i.e., eight female and eight male 'white' names and four female and four male 'black' names.[27] As mentioned, we present an overview of the different combinations of first and last names in Table A–1 in the Appendix. All names signalled the intended gender (male or female) and race (black or white). Additionally, all names were associated with the same social class (i.e., working or middle class), religious affiliation (i.e., Christianity, Atheism, or none in particular), and immigrant generation (i.e., this person is not an immigrant or third-generation or more).

[27] The results of the name categorization experiment are available upon request.

3. Data

3.1. Data collection process

We offered our online survey to participants via the online platform Prolific Academic (hereafter 'Prolific'). As mentioned, similar to the well-known platform Amazon Mechanical Turk (hereafter MTurk), Prolific is an online crowdsourcing platform on which scholars can recruit subjects to perform particular tasks in return for financial compensation (Palan & Schitter, 2018). An advantage of these online crowdsourcing platforms is that it permits us to recruit many participants in a relatively short amount of time.

Although MTurk has dominated the field for a long time, in recent years, Prolific has gained popularity in academics. Prolific is particularly attractive because it is designed to assist researchers with their academic studies. Moreover, Prolific has been found superior to other platforms, such as MTurk, in terms of data quality (Peer, Brandimarte, Samat, & Acquisti, 2017; Peer, Rothschild, Evernden, Gordon, & Damer, 2021). Not surprisingly, researchers from various disciplines, such as economics (e.g., Frimpong, Shuridah, Wilson, & Sarpong, 2020; Lucas, Berry, Giurge, & Chugh, 2021; Marreiros, Tonin, Vlassopoulos, & Schraefel, 2017; Singh, Crisafulli, Quamina, & Kottasz, 2020), psychology (e.g., Anwyl-Irvine, Massonnié, Flitton, Kirkhal, & Evershed, 2020; Callan, Kim, Gheorghiu, & Matthews, 2017; Costin & Vignoles, 2020; Dutt & Kohfeldt, 2019; Roster & Ferrari, 2020), and political sciences (e.g., Kaufmann, 2019; Peitz, Dhont, & Seyd, 2018), have used the platform to recruit participants for their studies.

As mentioned in Section 1, our target group of participants were U.S. individuals who had experience with assessing job candidates in the hiring process. To ensure only participants who met these conditions participated in our study, we used different pre-screening filters provided by Prolific. When individuals register as Prolific-users, they first have to fill out a general survey in which they are questioned on a broad range of subjects–among which different demographic characteristics (e.g., their nationality, education level, gender, and age) and work-related questions (e.g., whether they have any experience with hiring job applicants). The answers provided by Prolific-users in this general survey are then used as pre-screening filters for individual studies. By working with these pre-screening filters, researchers do not need to screen out possible participants at the beginning of their survey, therefore, eliminating the possibility that platform-users lie about their characteristics to qualify for a study and receive financial compensation.[28] In the context of this study, we used four pre-screening filters related

[28] An often mentioned criticism on the use of online crowdsourcing platforms (e.g., MTurk) to recruit participants is that platform-users are not necessarily representative for the group of people one aims to target because people can lie about their characteristics to qualify for a study. That this issue might occur has been proven by Sharpe Wessling, Huber, and Netzer (2017) who found that a considerable amount of MTurk-users (i.e., 20%) lies when filling in screening-questions at the beginning of a

to people's nationality, country of birth, country of residence, and experience in hiring.[29] Moreover, we included some additional screening questions at the beginning of our survey to check whether the filters of Prolific were still up to date and worked properly. Participants who provided answers to the screening questions that deviated from the pre-screening filters were redirected out of the survey.

For about one month (i.e., between the 6th of April and 8th of May 2021), Prolific-users could fill in the survey. Throughout the data collection, we subjected the completed surveys to a thorough quality control. That is, participants who filled out the survey in an extremely short amount of time, failed the attention check, or whose answers were clearly of low quality were excluded from our final sample. Moreover, participants with a high score on the Marlowe-Crowne Social Desirability Scale (MC-SDS) (i.e., higher than the mean score plus one standard deviation) and did not categorise the names of their assigned job applicants in the intended gender, race, social class, and religious categories in the manipulation check were excluded from our sample. In total, 403 participants filled in the survey completely and accurately. After taking into account participants' tendency to answer in a socially desirable way and their success in the name categorising exercise, 290 participants remained and were included in our final sample resulting in a total of about 1,450 observations, since the participants had to evaluate five vignettes each. We give a short description of the collected data in the next subsection.

3.2. Data description

In Table 3 (column 1), we present the descriptive statistics related to our total sample of participants (Panel A) and the twelve different job vacancies (Panel B). Moreover, to examine (i) whether the randomisation of the job applicants' gender over the participants had gone as planned and (ii) the job vacancies got evaluated with about the same frequency for each gender-group of job applicants, we divide our total sample of participants into two subgroups: (i) a subgroup who rated male job candidates (column 2) and (ii) a subgroup who rated female job applicants (column 3). We perform *T*-tests to check whether the different subgroups differed significantly from each other. X^2-tests yielded the same conclusions.

<Table 3 about here>

From column 1, it becomes clear that our sample matched our target sample of individuals: i.e., people

survey in order to qualify for a study and receive financial compensation. However, we believe that this criticism does not apply to platforms who work with pre-screening questions (e.g., Prolific) because platform-users can impossibly know based on which criteria future studies will select participants when filling in the general survey during registration on the platform.

[29] The screening questions consisted of answering 'United States' to the questions (i) 'What is your nationality?' (ii) 'What is your country of birth?', (iii) 'In what country do you currently reside?' and answering 'yes' to the question 'Do you have any experience in making hiring decisions (i.e. have you been responsible for hiring job candidates)?'

with experience in evaluating job applicants.[30] Indeed, 91.70% of our participants had at least one year of experience with assessing job candidates and 69.70% appraised job applicants, in general, at least once per semester in their current (or last) job. Moreover, most of our participants were full time employed (i.e., 81.00%) and had a university degree (i.e., 78.00%). Less than half of the participants were female (i.e., 43.50%), and about 62.10% was younger than 41. Furthermore, the majority identified themselves as 'white' (i.e., 82.80%), and 33.60% lived in the Southern region of the United States. Looking at Panel B of column 1, we also see that participants got assigned each of the twelve job vacancies with about the same frequency (except for the jobs of proof-reader and insurance sales agent–which were evaluated a little less often). When we focus on the descriptive statistics regarding the two subgroups of participants (i.e., columns 2, 3, and 4), we can see that the randomisation of our experimental stimuli over the participants was successful. Indeed, the participants of both subgroups are similar in terms of all presented participant characteristics. The same is true for the various job vacancies. Both male and female job candidates were evaluated for the different job vacancies with about the same frequency.

To check whether our sample of participants is representative of real-world recruiters in the United States, we compare some descriptive statistics of our sample with those of a sample of recruiters from the American Community Survey (ACS) in Table A–4 in the Appendix. We conducted different binomial tests (i.e., for the binary variables) and one one-sample T-test (i.e., to compare the mean age between the two samples) to examine whether the two samples significantly differed from each other. As becomes clear, our sample is not completely representative of real-world recruiters of the United States in terms of gender ($p = 0.000$) and age ($p = 0.000$). In terms of race-ethnicity and education level, the participants in our sample are, however, similar to those in the ACS ($p = 0.100$, $p = 0.289$, $p = 0.289$, respectively).

[30] All individuals who participated in our study had the American nationality and currently resided in the United States.

4. Results

4.1. Do the hiring chances of women vary by the gender prototypicality of a job applicant and the salience of gender in a particular context?

Although this is not the prime goal of our study, we first briefly examine whether we find hiring discrimination towards women in the context of our experiment.[31] Subsequently, we investigate whether we can identify some heterogeneity in women's hiring chances by the four factors discussed in Sections 1 and 2: i.e., the job applicants' race **(F1)** and family responsibilities **(F2)**–manipulated in our experiment to vary job applicants' gender prototypicality–the gender composition of the applicant pool **(F3)**, and the gender type of the job vacancies **(F4)**–manipulated in our experiment to vary the salience of gender in a particular hiring context.

To this end, we first run a linear baseline regression in which we regress the invitation probability on job applicants' gender and include the different other candidate characteristics (mentioned in Subsection 2.1 and presented in Table 1), the gender composition of the applicant pool, and the different job characteristics (discussed in Subsection 2.2 and summarised in Table A–2) as control variables (model 1).[32] In extended models, we add interaction terms between the applicants' gender and (i) the other candidate characteristics (model 2), the gender composition of the applicant pool (model 3), and the job characteristics (model 4) separately. In the last model (model 5), we include all interaction terms jointly. We apply linear regression models with standard errors clustered at the participant level.[33] We present the results of these analyses in Table 4.

<Table 4 about here>

Looking at column 1 (Table 4), we find that, overall, women are not treated differently from men in the context of our experiment. This finding is in line with many of the correspondence studies mentioned in Section 1. We do, however, find highly significant negative effects of multiple candidate characteristics that are, as such, not central to the present study on the probability to be invited to the second round of the selection process. That is, having (i) a large commuting distance (p = 0.000), (ii) no

[31] As mentioned in Section 2, vignette experiments are not the most appropriate research method to measure hiring discrimination due to its' laboratory setting–correspondence tests are.

[32] Note that we performed all statistical analyses in Stata and that the corresponding code is available upon request.

[33] We also ran ordered logistic regressions for the current and all following analyses, which yielded the same conclusions. In addition, we ran the abovementioned regressions (and all subsequent analyses) with an additional vector of participant characteristics. The results do not change upon this extended specification. We decided to present the results of the analyses without the participant characteristics because (i) the coefficients regarding these characteristics are not causally interpretable (since we did not manipulate them experimentally in our study), (ii) they form no source of bias (see Subsection 3.2 where we point out that the randomisation of our experimental stimuli over the participants was successful), and (iii) for conciseness.

experience in the occupation (p = 0.000), or (iii) an employment gap in one's employment history (p = 0.000) negatively influences job candidate's hiring chances. In contrast, we also find a significant positive effect of volunteering on one's chances to be invited to the second round (p = 0.034).

Next, if we focus on possible moderation effects (Table 4, columns 2, 3, 4, and 5), we find no heterogeneity in women's hiring chances by both job applicants' race **(F1)** and family responsibilities **(F2)**. The same is true for the interactions between the candidates' gender and the other candidate characteristics, as well as for the interaction terms between the job applicants' gender and the various job characteristics. In contrast to some of the existing correspondence tests mentioned in Section 1, we, therefore, do not find that women's hiring chances vary by the gender type of the occupation **(F4)** nor the gender prototypicality of the subgroup of women considered.

However, we do find a (weakly) significant positive moderation effect of the gender composition of the applicant pool **(F3)** on women's chances to be invited to the second round of the selection process. In other words, we find that women's hiring chances are higher when the applicant pool consists mainly out of men. When we focus on model 4 (i.e., Table 4, column 4), we see that this effect is only significant on the 10%-level (p = 0.061). However, when we consider model 5 (i.e., the model in which we include all interaction terms jointly), it becomes significant at the 5%-level (p = 0.039).

As a robustness check, we reran the different regression analyses mentioned above with a sample of participants with a very low score on the social desirability scale (i.e., a score lower than the sample mean of 0.440; N = 875). Also when we restrict our sample to those participants who are not inclined to answer in a socially desirable way, we find similar results. Detailed regression results can be obtained upon request.

4.2. Are gender stigma present in the hiring context?

Although we do not find a gender effect on job applicants' hiring chances overall in the context of our experiment, this does not necessarily mean that no gender stigmas towards women exist. Indeed, it might be possible that both positive and negative gender stereotypes about women are present but weigh each other out statistically. Therefore, in the current subsection, we first investigate whether the gender stereotypes existing about men and women in society are also identifiable in the hiring context (i.e., **H1**).

To this end, we run 21 regressions in which we regress the scales related to the different gender stereotypes discussed in Subsection 2.2 (and summarised in Table 2) on candidates' gender. Again, we include the other candidate characteristics, the gender composition of the applicant pool, and the various job characteristics as control variables and correct the standard errors for clustering of the observations at the participant level. We present the results of these regression analyses in Table 5.

<Table 5 about here>

As becomes clear from Table 5, we find that the participants in our experiment to some extent do perceive women in a stereotypical way. More concretely, in line with previous research, we find that participants perceive women to have more social abilities ($p = 0.012$) and to be more supportive ($p = 0.006$), creative ($p = 0.021$), and open to new people and experiences ($p = 0.028$) compared with men—i.e., in line with our expectations, perceptions that correspond to the idea that women are more communal and have better soft skills than men. In line with the idea that women are less agentic than men, we also find that participants perceive the female job applicants in our experiment to be less assertive ($p = 0.050$). Other stereotypes about women that are valent in our study are the perception that women are more absent due to childcare responsibilities ($p = 0.000$) and have lower physical abilities ($p = 0.004$). Concerning the willingness of participants (and co-workers and customers) to cooperate with a job applicant, we find a greater willingness to collaborate both by the participants themselves ($p = 0.010$) as by their co-workers ($p = 0.027$) and customer ($p = 0.019$) (as believed by employers) if the job candidate is a woman.

If we, however, focus on the sample of participants with a very low score on the MC-SDS, the results slightly change. That is, the positive perceptions that women are more sociable, more supportive, more creative, and more open become less significant–or in one case even insignificant–(i.e., $p = 0.036$, $p = 0.068$, $p = 0.052$, $p = 0.125$, respectively), while the negative perceptions that women are less assertive, have less physical abilities, and fall out more due to family responsibilities stay (highly) significant (i.e., at the 5%- or the 1%-significance level). Moreover, the higher willingness of participants, co-workers, or customers (as believed by the participants) to collaborate with female job applicants compared with male job candidates disappears completely. Therefore, our results seem to suffer to some extent from social desirability bias. Nevertheless, some stereotypical perceptions remain present in the context of our experiment, even when we consider social desirability bias. As a result, we can claim that we cannot reject **(H1)**.

Taken together, the fact that we do not find any hiring discrimination towards women overall in the context of our experiment can thus be explained by the fact that the negative employer perceptions of women identified above are somewhat compensated by the positive employer perceptions and, therefore, do not lead to lower (or higher) hiring chances for women compared with men.

Moreover, although these are not the focus of our study, we also find significant effects of other candidate characteristics on different employer perceptions. We discuss some of them briefly in the following paragraph. More concretely, we find that black job applicants are perceived to be more motivated (p = 0.040) and to have better social skills than white job applicants. However, this last effect is only significant on the 10%-level (p = 0.083). Additionally, we find that mentioning an employment gap due to family responsibilities negatively affects employers' perceptions of job applicants' (i) intellectual abilities, (ii) physical abilities, (iii) ambition, (iv) motivation, (v) autonomy, (vi) resilience, (vii) reliability, (viii) cooperativeness, (ix) flexibility, (x) openness, and (xi) sense of responsibility. All these effects are significant at the 5%-level at least. Besides this, employers also believe that those job applicants who mention an employment gap due to caretaking responsibilities will also be more absent in the future due to caretaker duties (p = 0.000). Furthermore, having no experience in the occupation or a large commuting distance affects employer's perceptions negatively regarding (almost) all statements. The same can be said for job applicants with an employment gap due to unemployment or health issues. Only employers' perception regarding a job applicant's absence due to childcare responsibilities (respect towards authority) does not differ between job applicant's that have been unemployed (have had health issues) and those who did not report an employment gap in their résumé. Lastly, while mentioning an employment gap has almost exclusively negative effects on employers' perceptions of job applicants, the contrary is true for mentioning an extra-curricular activity on one's résumé—especially when this extra-curricular activity implies volunteer work or sports. Again, the results slightly change when we rerun our analyses with a subsample of participants with a score on the MC-SDS below 0.440. However, because these effects are not the focus of our study, we decide not to discuss them here. The results of these robustness analyses can be obtained on request.

4.3. Do these stigma vary by the gender prototypicality of a job applicant and the salience of gender in a particular context?

After identifying gender stereotypes in the hiring process about women in general, in a second step, we now examine whether the degree to which these stereotypes are triggered varies by the four factors discussed in Sections 1 and 2. Indeed, as argued by Ridgeway and Kricheli-Katz (2013), it is not because particular gender stigmas do or do not occur in general that they also do or do not occur in specific situations. As mentioned in Section 1, we hypothesised that the stereotypical perceptions about women will be triggered more heavily in male gender-typed jobs **(H2)** and male-dominated applicant pools **(H3)**. Moreover, we hypothesised that the stereotypes about women are triggered less when the female job applicant is black **(H4)** but more when the female job applicant explicitly mentions family responsibilities on her résumé **(H5)**.

To test the abovementioned hypotheses, we again run the same 21 regressions as in Table 5, however, extending them by jointly including interaction terms between the job applicant's gender and the different (i) other candidate characteristics—most importantly: someone's race and employment gap due to family responsibilities—(ii) the gender composition of the applicant pool, and (iii) the different job characteristics to the regression. We present the results of these regression analyses in Table 6.

<Table 6 about here>

From Table 6, we can see that some of the stereotypes about women in general are, as hypothesised, to a lesser extent applicable to African American women. That is, the idea that women would be more supportive, more cooperative, and more creative (i.e., all stereotypes identified in the previous step or former studies) are to a lesser extent triggered when the female job applicant is black (p = 0.006, p = 0.012, and p = 0.032, respectively). Moreover, the positive stereotypes about women's social abilities, openness towards new people and experiences, and respect towards authority are, in line with what we hypothesised, less triggered if the job applicant is an African American woman. These effects are, however, only significant at the 10%-level. Furthermore, although we did not find any evidence for the existence of the negative gender stereotypes on women's intellectual abilities and accuracy identified in the literature for all women in our experiment, we do find that these negative stereotypes are triggered when the job applicant is an African American woman (p = 0.040 and p = 0.001, respectively). Although these results seem unexpected at first sight—i.e., we expected the stereotypes about traditional women to be triggered less for unprototypical women—a potential explanation for these results can be found in the stereotypes existing about black people. Indeed, previous research has found that blacks–in general–are perceived to be less competent (e.g., Federal Glass Ceiling Commission,1995; Ghavami & Peplau, 2013; Zou & Cheryan, 2017) and less accurate compared with whites (e.g., Kirschenman & Neckerman,1991). Consequentially, because African American women are both black and female, the negative stereotypes about women, in general, could be reinforced by the negative stereotypes existing about blacks in U.S. society. These results are, therefore, very much in line with the notion of intersectionality coined by Crenshaw (1989).[34, 35]

Again, as a robustness check, we reran the abovementioned regression analyses with a subsample of participants with a very low score on the MC-SDS. The results remain more or less the same when we focus on this subsample except for the penalty that women face regarding their perceived assertiveness. The belief that women would be less assertive than men seems to be less pronounced when the female

[34] Although the term 'intersectionality is attributed to Crenshaw (1989), its theoretical framework already arose earlier in work by African American feminists (e.g., Frances Beal, 1970; Bell Hooks (1984), and Toni Cade Bambara (1970)).

[35] Rosette et al. (2018) define intersectionality as "overlapping social categories, such as race and gender, that are relevant to a specified individual or groups' identity and create a unique experience that is separate and apart from its originating categories.

job applicant is black ($p = 0.027$). Altogether, these results prove that the stereotypes about women—more generally—are not applied to the same degree to black women as to white women. In other words, the non-prototypicality of African American women, indeed, seems to affect how they are perceived (in terms of gender stereotypes) in the hiring context. For most stereotypes identified, we found results in line with our hypothesis **(H4)**, i.e., the general gender stereotypes about women are triggered less if the job applicant is black. However, we did find that some (negative) stereotypes about women were more pronounced (i.e., the stereotypes about their intellectual abilities and accuracy) when the female job applicant was black–and thus less prototypical for the female gender. Therefore, our fifth hypothesis does not entirely hold.

Note also that these varying perceptions towards black and white women do not seem to translate to different hiring chances between these two subgroups of women in the context of our experiment. A potential explanation for this is the research method we applied. Indeed, as mentioned in Section 2, vignette experiments are less appropriate to measure women's hiring chances objectively than correspondence tests.

If we then focus on women who mention many caretaker responsibilities in their résumés, the results are somewhat less in line with what we expected based on the theoretical framework of Ridgeway and Kricheli-Katz (2013). As mentioned in Section 1, we expected women who mention family responsibilities on their résumés to be perceived more in line with traditional gender stereotypes about women because they are more prototypical for their gender category **(H5)**. Although this is the case for some of the surveyed stereotypes—i.e., the negative stereotypes about women's (i) intellectual abilities, (ii) physical abilities, (iii) assertiveness, (iv) motivation, (v) resilience, (vi) flexibility, and (vii) accuracy are, indeed, triggered more heavily when women mention caretaking responsibilities on their résumé—this is not true for others.[36] More concretely, in contrast to what we expected, the positive stereotypes that women would be (i) more social, (ii) more supportive, (iii) more open, (iv) more cooperative, and (v) have more respect towards authority are triggered less if the female job applicant mentions caretaker responsibilities on their résumé. Moreover, the greater willingness of participants (and co-workers and customers as perceived by the participants) to collaborate with a female job applicant (found in Table 5) is also less pronounced when the female job applicants mention an employment gap due to family responsibilities on their résumé. Furthermore, also in stark contrast to what we expected, the negative stereotype that women will fall out of the job due to childcare

[36] In line with our expectations, we also find that the negative gender stereotypes about women's reliability and sense of responsibility are triggered more heavily when women mention caretaking responsibilities on their resumes. However, these moderation effects are only weakly significant (i.e., significant at the 10%-level). We, therefore, cannot rule out that these results are obtained by coincidence.

responsibilities is less present for women who have been inactive due to family responsibilities in the past. That is, women with a period of inactivity due to family responsibilities seem to be penalised less than other women in terms of future absenteeism due to caretaking duties. However, this finding is more intuitive when we rephrase it: a career gap due to family responsibilities triggers the signal of a future dropout due to childcare more strongly for men. Indeed, it seems logical to believe that the signal of future dropout due to childcare is more strongly triggered for men with an employment gap due to family responsibilities than for women with such an employment gap because it is less common for men to be inactive for some time due to caretaking responsibilities than it is for women.

If we then rerun the analyses with a subsample of participants with a very low score on the MC-SDS, we find more or less similar results–albeit that the results, overall, are less significant than before (i.e., on the 5% or 10% significance level)–except for the stereotypes concerning women's resilience and supportiveness. Where in the full sample, these stereotypes were more (for the resilience-stereotype) or less (for the supportiveness-stereotype) pronounced when the female job applicant mentioned caretaker responsibilities, in the restricted sample, these moderation effects are insignificant.

The positive stereotypes about women are, thus, triggered less when a female job applicant mentions family responsibilities in her résumé, whereas the negative stereotypes are triggered more, except for the negative stereotype concerning absenteeism. The latter channel, therefore, seems to be important, given that the overall negative effect of family responsibilities on someone's chances to be invited to the second round is not more pronounced for female candidates (see Table 4).

Although we did find support for our prototypicality hypothesis if we consider African American women, the same cannot be said convincingly if we take into account women's caretaker responsibilities as a signal for prototypicality. A potential explanation for the fact that women with caretaking responsibilities are overall rated more negatively compared with other job applicants in the hiring context is the idea that the culturally internalised beliefs about the 'ideal worker' are incompatible with women's caretaker role, leading women with caretaking responsibilities to be perceived as 'bad workers' overall (Correll, Bernard, & Paik, 2007).

Next, if we then focus on the effect of the salience of someone's gender in a particular hiring context on the degree to which gender stereotypes are triggered and potentially influence employers hiring decisions, we find that the gender composition of the applicant pool does not seem to affect the extent to which stereotypes about women are triggered. The same is true when we rerun our analyses with a subsample of participants with a (very) low tendency to socially desirable answering. Therefore, we reject our third hypothesis **(H3)**. Moreover, we find that also the gender type of a job influences the degree to which gender stereotypes about women are triggered to a minor extent. Indeed, we only find two significant moderation effects related to applying for a male-dominated job. First, the stereotype that women are less assertive is more pronounced when they apply for male-dominated jobs (p = 0.050). Second, also their accuracy is less favourably evaluated in these kinds of jobs (p = 0.032). These results are in line with our expectations based on the theoretical framework of Ridgeway and Kricheli-Katz (2013)—i.e., that the stereotypes about women will be triggered more heavily in male gender-typed jobs **(H2)**. However, when we rerun the analyses with a subsample of participants with a score on the MC-SDS below 0.440 (i.e., the sample mean), both effects become insignificant. We, therefore, cannot rule out that these results are somewhat biased due to social desirability.

Altogether, our results indicate that the prototypicality of a female job applicant in terms of gender seem to influence the degree to which gender stereotypes are triggered–and potentially also affect the hiring chances of women–more than the gender composition of an applicant pool or the gender type of a job. Our results, therefore seem to confirm our application of the theory developed by Ridgeway and Kricheli-Katz (2013) only partially.

5. Conclusion

To investigate potential explanations for the varying hiring chances of women identified in former correspondence tests, we conducted an innovative vignette experiment in the United States with 290 individuals with genuine experience in hiring. These participants were asked to evaluate multiple fictitious job applicants regarding their hireability and 21 statements related to gender stereotypes drawn from a systemic review of the literature. Moreover, we experimentally manipulated both the prototypicality of job applicants in terms of gender (manipulated via the race and family responsibilities of a job applicant) and the salience of gender in a particular hiring context (manipulated via the gender composition of the applicant pool and the gender type of a job) in our experiment. In doing so, we could investigate (i) how women are perceived by employers in the hiring context and (ii) how these perceptions vary by the gender prototypicality of a job applicant and the salience of gender in a particular context, two key dimensions in the theoretical framework developed by Ridgeway and Kricheli-Katz (2013).

We found that employers perceived women to be more social, supportive, open, and creative than men. Moreover, we found that women were perceived to have fewer physical abilities and to be less assertive than men, but also more absent due to childcare responsibilities. These perceptions seem to cancel each other out statistically since we do not find different hiring chances between men and women overall.

Furthermore, our results indicate that the gender prototypicality of job applicants influence employers' perceptions about women more than the gender composition of the applicant pool or the gender type of a job. Indeed, employer perceptions about female job applicants changed if the job applicant was black or mentioned an unemployment gap due to family responsibilities in her résumé. That is, first, the negative stereotypes about women's physical abilities and assertiveness are, indeed, triggered more heavily when women mention caretaking responsibilities on their résumé. In other words, also the potential caretaker role of women influences employers' perceptions towards them. Second, the ideas that women are more supportive and more creative are to a lesser extent triggered when the female job applicant is black. Moreover, although we did not find any evidence for the existence of the negative gender stereotypes on women's intellectual abilities and accuracy identified in the literature for all women in our experiment, we do find that these negative stereotypes are triggered when the female job applicant is black. In the context of our experiment, African American women, thus, are perceived more negatively than white women. Although these more negative perceptions towards black women do not translate to different hiring chances between white and black women in the context of our experiment, this does not mean they may not influence other work-related decisions (e.g., promotion, remuneration, or dismissal decisions). It could, thus, be interesting for future research to apply and test the theoretical framework of Ridgeway and Kricheli-Katz (2013) in these other contexts (as well as contexts outside the labour market). We also find some suggestive evidence that both the stereotypes existing about women and blacks might reinforce (or weaken—depending on the content of the stereotypes) each other. It would, therefore, be interesting for future research to use a more intersectional approach and investigate the interplay between gender and race stereotypes existing in society while taking into account both the prototypicality of job applicants (in terms of gender and race) and the salience of gender or race in a particular hiring context.

We contribute to the existing literature by being the first to empirically test the theoretical framework of Ridgeway and Kricheli-Katz (2013) applied to the hiring context to study potential explanations for the varying hiring chances of women, therefore, introducing a new and promising theoretical framework in the field of economics. Indeed, as mentioned in Section 1, economists often appeal to Beckers' (1957) model of taste-based discrimination or Arrow (1973) and Phelps (1972) model of statistical discrimination to explain hiring discrimination towards women. However, these models are limited in the sense that they treat either employers' tastes or stereotypical beliefs as constant over social situations. By introducing another theoretical framework in the field of economics, we hope to inspire other economists to consider theories beyond the models of taste-based (Becker, 1957) and statistical discrimination (Arrow, 1973; Phelps, 1972) to gain insights into the underlying mechanisms and moderators of hiring discrimination–towards women, but also towards other disadvantaged groups in society. Moreover, we contribute methodologically by testing the names used in our experiment in a preliminary name categorising experiment, therefore taking into account Gaddis (2017, 2019) justified criticism that the names used in former experiments are often noisy signals for race (or gender).

Next to its academic relevance, our study is also interesting from a policy perspective. Knowing how different women are perceived in particular situations in the hiring process is crucial information to develop targeted policy measures that counter barriers experienced by women in the labour market. Indeed, the fact that women with caretaking responsibilities are viewed more negatively by employers—together with the fact that today many individuals experience some lapse in employment due to care for children or other family members (Weisshaar, 2018)—calls for policy measures that, especially, support women with caretaker responsibilities. The same can be said about policy measures targeted to aid African American women. Moreover, our results are also valuable for women seeking a job because it allows them to proactively counter negative stereotypes existing about them by providing additional information to employers that may debunk these negative stereotypes. For example, women–especially women with an employment gap due to family responsibilities–might benefit from signalling sufficient levels of assertiveness in their résumés.

Although our research greatly contributes to the existing literature, our vignette experiment design does not come without limitations. First, our research is limited by its laboratory setting. In lab experiments, participants are aware they are participating in an experiment and that their answers have no real-world consequences. Although a lab experiment is advantageous from a research-ethical point of view (Charness, Gneezy, & Kuhn, 2013; Riach & Rich, 2004) and essential for obtaining deeper insights into thought processes (Baert & De Pauw, 2014; Van Belle et al., 2018; Van Hoye & Lievens, 2003), it could induce a certain degree of hypothetical bias. That is, participants might behave differently in our experiment than in real life, for example, because they do not take the survey seriously or try to hide the fact that they are inclined to discriminate (i.e., social desirability bias).

That hypothetical bias can be a problem in factorial survey experiments is proven by the study of Pager & Quillian (2005). In their study, the authors conducted an audit experiment and a follow-up telephone survey with the same employers to show that employers' hiring intentions identified via the experimental survey did not always reflect their actual hiring decisions measured in the audit experiment. However, other research validating results of survey experiments against real-world behaviour measured in field experiments (outside the field of labour market research) found similar impacts of the characteristics of interests on both participants' behavioural intentions and corresponding real-life behaviour (e.g., Hainmueller, Hangartner, & Yamamoto, 2015; Nisic & Auspurg, 2009). The results regarding the external validity of factorial survey experiments are, thus, mixed.

In our study, we aimed to minimise this hypothetical bias in three ways. As mentioned in Section 2, we designed our experiment to mimic real-life hiring decisions as closely as possible to make sure the participants in our experiment did not behave very differently than in real life. By simultaneously manipulating different applicant characteristics, we aimed to imitate the complex nature of hiring decisions in the field, where HR managers and employers are also confronted with the evaluation of job applicants differing in several personal characteristics such as gender, educational level, and work experience (Baert & De Pauw, 2014; Colquitt, 2008; Shadish, Cook, & Campbell, 2002). Additionally, to assure the participants in our sample took the survey seriously, we screened all completed surveys on the quality of the data and retained only the participants who filled out the survey completely and accurately (see Subsection 3.1). For example, participants who filled out the survey in an extremely short amount of time, failed the attention check(s), or whose answers were clearly of a low quality (i.e., all indications that the participants did not take the survey seriously) were excluded from our final sample. Lastly, we controlled for socially desirable answering by including a social desirability scale in our survey and running the analyses only for the subset of participants with a low score on this scale.

Although, we designed our experiment as to minimise hypothetical bias, we cannot claim that the results obtained with this factorial survey experiment, indeed, coincide with employers actual real-life hiring behaviour. Our results should, therefore, be interpreted as employers' hiring intentions in a simplified context and not as real-life hiring behaviour.

Second, although we aimed to increase the external validity of our research by recruiting a large sample of people with genuine experience in the hiring context via Prolific, our sample is not entirely representative of real-life HR professionals (see Subsection 3.2). More concretely, our sample contains more men and is also, on average, younger than the representative sample of HR managers from the ACS. Therefore, we believe it would be interesting for future research to replicate the study with a representative sample of HR professionals.

Third, we cannot generalise our results to other subgroups of women or contexts. Indeed, as demonstrated in this study, the stereotypes and attitudes about different subgroups of women (i.e., African American women and women with caretaker responsibilities) differ considerably. Therefore, we believe it would also be interesting to investigate the stereotypes existing about other subgroups of women by manipulating the prototypicality of women in another way–for example, by focussing on women with a different sexual orientation or gender identity. Moreover, although we decided to focus on the hiring context in this study, the broad and general theoretical framework of Ridgeway and Kricheli-Katz (2013) might also easily be translated to other contexts, such as the promotion or dismissal contexts–or even contexts outside the labour market. Therefore, we encourage researchers to conduct this study in different contexts as well.

References

Acquisti, A., & Fong, C. (2020). An experiment in hiring discrimination via online social networks. *Management Science, 66*(3), 1005–1024.

Albert, R., Escot, L., & Fernández-Cornejo, J. A. (2011). A field experiment to study sex and age discrimination in the Madrid labour market. *The International Journal of Human Resource Management*, *22*(2), 351–375.

Ambuehl, S., & Ockenfels, A. (2017). The ethics of incentivizing the uninformed: A vignette study. *American Economic Review*, *107*(5), 91–95.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods, 52*(1), 388–407.

Arrow, K. J. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in labor markets.* Princeton, NJ: Princeton University Press.

Asali, M., Pignatti, N., & Skhirtladze, S. (2018). Employment discrimination in a former Soviet Union Republic: Evidence from a field experiment. *Journal of Comparative Economics*, *46*(4), 1294–1309.

Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology, 6*(3), 128–138.

Auspurg, K., & Hinz, T. (2014). *Factorial survey experiments*. Thousand Oaks, CA: Sage.

Auspurg, K., Hinz, T., & Sauer, C. (2017). Why should women get less? Evidence on the gender pay gap from multifactorial survey experiments. *American Sociological Review, 82*(1), 179–210.

Baert, S. (2018a). Hiring discrimination: An overview of (almost) all correspondence experiments since 2005. In S. Gaddis (Ed.), *Audit studies: Behind the scenes with theory, method, and nuance* (4^{th} ed.). Cham: Springer.

Baert, S. (2018b). Hiring a gay man, taking a risk? A lab experiment on employment discrimination and risk aversion. *Journal of Homosexuality, 65*(8), 1015–1031.

Baert, S., De Meyer, A.S., Moerman, Y., & Omey, E. (2018). Does size matter? Hiring discrimination and firm size. *International Journal of Manpower*, *39*(1), 550–566.

Baert, S., & De Pauw, A. S. (2014). Is ethnic discrimination due to distaste or statistics?. *Economics Letters, 125*(2), 270–273.

Baert, S., De Pauw, A. S., & Deschacht, N. (2016). Do employer preferences contribute to sticky floors?. *ILR Review*, *69*(3), 714–736.

Bambara, T. C. (1970). *The black woman: An anthology*. New York, NY: New American Library.

Beal, F. M. (2008). Double jeopardy: To be Black and female. *Meridians, 8*(2), 166–176.

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

Benhabib, L., & Adair, P. (2017, June 29). *Hiring Discrimination on the Algerian Labour Market: an Assessment with Testing* [Conference presentation]. Africa Meeting of the Econometric Society, Algiers, Algeria.

Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne social desirability scale. *Educational and Psychological Measurements, 62*(4), 570–589.

Berger, J. & Webster, M. (2006). Expectations, status, and behavior. In P. J. Burke (Ed.), *Contemporary social psychological theories*. Stanford: Stanford University Press.

Berson, C. (2012). *Does Competition Induce Hiring Equity?* (CES Working Paper No. 19). Centre d'Economie de la Sorbonne.

Bertogg, A., Imdorf, C., Hyggen, C., Parsanoglou, D., & Stoilova, R. (2020). Gender Discrimination in the Hiring of Skilled Professionals in Two Male-Dominated Occupational Fields: A Factorial Survey Experiment with Real-World Vacancies and Recruiters in Four European Countries. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *72*(1), 261–289.

Brandén, M., Bygren, M., & Gähler, M. (2018). Can the trailing spouse phenomenon be explained by employer recruitment choices? *Population, Space and Place, 24*(6), 1–11.

Bright, J., & Davies, F. (1999). You Did What in 1983? The Effects of Explained and Unexplained Gaps in Career History Presented on Résumés. *Australian Journal of Career Development, 8*(1), 12–17.

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

Bosak, J., & Sczesny, S. (2011). Gender bias in leader selection? Evidence from a hiring simulation study. *Sex roles*, *65*(3), 234–242.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-Role Stereotypes: A Current Appraisal. *Journal of Social issues, 28*(2), 59–78.

Busetta, G., Campolo, M. G., & Panarello, D. (2018). Immigrants and Italian labor market: statistical or taste-based discrimination? *Genus, 74*(1), 1–20.

Bygren, M., Erlandsson, A., & Gähler, M. (2017). Do employers prefer fathers? Evidence from a field experiment testing the gender by parenthood interaction effect on callbacks to job applications. *European Sociological Review, 33*(3), 337–348.

Callan, M. J., Kim, H., Gheorghiu, A. I., & Matthews, W. J. (2017). The interrelations between social class, personal relative deprivation, and prosociality. *Social Psychological and Personality Science, 8*(6), 660–669.

Capéau, B., Eeman, L., Groenez, S., & Lamberts, M. (2012). *Two concepts of discrimination: inequality of opportunity versus unequal treatment of equals* (Ecore Discussion Papers No. 58). The International Association for Research and Teaching.

Carlsson, M. (2010). Experimental evidence of discrimination in the hiring of first-and second-generation immigrants. *Labour, 24*(3), 263–278.

Carlsson, M. (2011). Does hiring discrimination cause gender segregation in the Swedish labor market? *Feminist Economics, 17*(3), 71–102.

Carlsson, M., & Eriksson, S. (2019). Age discrimination in hiring decisions: Evidence from a field experiment in the labor market. *Labour Economics*, *59*, 173–183.

Carlsson, M., Reshid, A. A., & Rooth, D. O. (2018). Neighborhood signaling effects, commuting time, and employment: Evidence from a field experiment. *International Journal of Manpower, 39*(4), 534–549.

Carolina, A. (2005). Relational demography within groups: Through the lens of discrimination. In R. L. Dipboye & A. Collela (Eds.) *Discrimination at work: The psychological and organizational bases*. Hove: Psychology Press.

Charness, G., Gneezy, U., & Kuhn, M. A. (2013). Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics. *Journal of Economic Behavior & Organization, 91*, 93–100.

Clow, K. A., Ricciardelli, R., & Bartfay, W. J. (2015). Are you man enough to be a nurse? The impact of ambivalent sexism and role congruity on perceptions of men and women in nursing advertisements. *Sex roles, 72*(7), 363–376.

Collins, P. H. (2000). *Black feminist thought*. New York, NY: Routledge.

Collins, P. H. (2004). *Black sexual politics: African Americans, gender, and the new racism*. New York, NY: Routledge.

Colquitt, J.A. (2008). From the editors publishing laboratory research in AMJ: A question of when, not if. *Academy of Management Journal, 51*(2008), 616–620.

Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology, 112*(5), 1297–1338.

Costin, V., & Vignoles, V. L. (2020). Meaning is about mattering: Evaluating coherence, purpose, and existential mattering as precursors of meaning in life judgments. *Journal of Personality and Social Psychology, 118*(4), 864–884.

Crenshaw, K. W. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine. In *University of Chicago Legal Forum* (pp. 139–168). Chicago, IL: University of Chicago Law School.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social issues, 60*(4), 701–718.

Cuddy, A. J., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in organizational behavior, 31*, 73–98.

Cunningham, G. B., Sartore, M. L., & McCullough, B. P. (2010). The influence of applicant sexual orientation, applicant gender, and rater gender on ascribed attributions and hiring recommendations of personal trainers. *Journal of Sport Management*, *24*(4), 400–415.

Damelang, A., & Abraham, M. (2016). You Can Take Some of It with You! A Vignette Study on the Acceptance of Foreign Vocational Certificates and Ethnic Inequality in the German Labor Market. *Zeitschrift für Soziologie, 45*(2), 91–106.

Damelang, A., Abraham, M., Ebensperger, S., & Stumpf, F. (2019). The hiring prospects of foreign-educated immigrants: A factorial survey among German employers. *Work, Employment and Society*, *33*(5), 739–758.

Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2016). Race and gender effects on employer interest in job applicants: new evidence from a résumé field experiment. *Applied Economics Letters, 23*(12), 853–856.

Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior*, *56*(2), 225–248.

Derous, E., Nguyen, H. H., & Ryan, A. M. (2009). Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance, 22*(4), 297–320.

Derous, E., Ryan, A. M., & Nguyen, H. H. D. (2012). Multiple categorization in résumé screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings. *Journal of Organizational Behavior, 33*(4), 544–570.

Di Stasio, V. (2014). Education as a signal of trainability: Results from a vignette study with Italian employers. *European Sociological Review, 30*(6), 796–809.

Di Stasio, V., Lancee, B., Veit, S., & Yemane, R. (2021). Muslim by default or religious discrimination? Results from a cross-national field experiment on hiring discrimination. *Journal of Ethnic and Migration Studies, 47*(6), 1305–1326.

Dipboye, R. L., Arvey, R. D., & Terpstra, D. E. (1977). Sex and physical attractiveness of raters and applicants as determinants of résumé evaluations. *Journal of Applied Psychology*, *62*(3), 288.

Donovan, R. A. (2011). Tough or Tender: (Dis) Similarities in White College Students' Perceptions of Black and White Women. *Psychology of Women Quarterly, 35*(3), 458–468.

Drydakis, N. (2010). Religious affiliation and employment bias in the labor market. *Journal for the Scientific Study of Religion, 49*(3), 477–493.

Drydakis, N., Sidiropoulou, K., Bozani, V., Selmanovic, S., & Patnaik, S. (2018). Masculine vs feminine personality traits and women's employment outcomes in Britain: a field experiment. *International Journal of Manpower, 39*(4), 621–630.

Duguet, E., Parquet, L. D., L'Horty, Y., & Petit, P. (2015). New Evidence of Ethnic and Gender discriminations in the French Labor Market using experimental data: A ranking extension of responses from correspondence tests. *Annals of Economics and Statistics/Annales d'Économie et de Statistique, (117/118),* 21–39.

Duguet, E. & Petit, P. (2005). Hiring discrimination in the French financial sector: an econometric analysis on field experiment data. *Annals of Economics and Statistics/Annales d'Economie et de Statistique*, 78, 79-102.

Dülmer, H. (2007). Experimental plans in factorial surveys: random or quota design? *Sociological Methods & Research, 35*(3), 382–409.

Dutt, A., & Kohfeldt, D. (2019). Assessing the relationship between neoliberal ideology and reactions to Central American asylum seekers in the United States. *Journal of Social Issues, 75*(1), 134–152.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, *109*(3), 573.

Edwards, G. H. (1992). The structure and content of the male gender role stereotype: An exploration of subtypes. *Sex roles, 27*(9-10), 533–551.

Eriksson, T., & Kristensen, N. (2014). Wages or fringes? Some evidence on trade-offs and sorting. *Journal of Labor Economics, 32*(4), 899–928.

Eriksson, S., & Rooth, D. O. (2014). Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. *American economic review, 104*(3), 1014–39.

Federal Glass Ceiling Commission. (1995). *A Solid Investment: Making Full Use of the Nation's Human Capital: Recommendations of the Federal Glass Ceiling Commission*. The Commission.

Fiske, S.T., Cuddy, J.A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902.

Firth, M. (1982). Sex discrimination in job opportunities for women. *Sex Roles, 8*(8), 891–901.

Foschi, M., & Valenzuela, J. (2008). Selecting job applicants: Effects from gender, self-presentation, and decision type. *Social Science Research*, *37*(3), 1022–1038.

Frimpong, K., Shuridah, O., Wilson, A., & Sarpong, F. (2020). A cross-national investigation of trait antecedents of mobile-banking adoption. *Thunderbird International Business Review, 62*(4), 411–424.

Gaddis, S. M. (2017). How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies. *Sociological Science, 4*, 469–489.

Gaddis, S. M. (2019). *Signaling class: An experiment examining social class perceptions from names used in correspondence audit studies* (SSRN Working Paper No. 3350739). Social Science Research Network.

Galarza, F. B., & Yamada, G. (2019). Afro-descendants in Peru: Do beauty and race matter in the labor market? *Review of Development Economics*, *23*(1), 211–230.

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological science, 24*(4), 498–506.

Ghavami, N., & Peplau, L. A. (2013). An intersectional analysis of gender and ethnic stereotypes: Testing three hypotheses. *Psychology of Women Quarterly, 37*(1), 113–127.

Goff, P. A., Thomas, M. A., & Jackson, M. C. (2008). "Ain't I a woman?": Towards an intersectional approach to person perception and group-based harms. *Sex Roles*, *59*(5), 392–403.

González, M. J., Cortina, C., & Rodríguez, J. (2019). The role of gender stereotypes in hiring: A field experiment. *European Sociological Review, 35*(2), 187–204.

Guryan, J., & Charles, K. K. (2013). Taste-based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots. *The Economic Journal, 123*(572), 417–432.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences, 112*(8), 2395–2400.

Heilman, M. E. (1980). The impact of situational factors on personnel decisions concerning women: Varying the sex composition of the applicant pool. *Organizational Behavior and Human Performance, 26*(3), 386–395.

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in organizational behavior*, *5,* 269–298.

Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior and Personality, 10*(4), 3–26.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of social issues*, *57*(4), 657–674.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in organizational Behavior, 32*, 113–135.

Hooks, B. (2015). Black women: shaping feminist theory. *Revista Brasileira de Ciência Política*, (16), 193–210.

Hosoda, M., Stone, D. L., & Stone-Romero, E. F. (2003). The Interactive Effects of Race, Gender, and Job Type on Job Suitability Ratings and Selection Decisions. *Journal of Applied Social Psychology, 33*(1), 145–178.

Jacquemet, N., & Yannelis, C. (2012). Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market. *Labour Economics, 19*(6), 824–832.

Jasso, G. (2006). Factorial survey methods for studying beliefs and judgments. *Sociological Methods and Research, 34*(3), 334–423.

Kaufmann, E. (2019). Ethno-traditional nationalism and the challenge of immigration. *Nations and Nationalism, 25*(2), 435–448.

Keuschnigg, M., & Wolbring, T. (2016). The use of field experiments to study mechanisms of discrimination. *Analyse & Kritik*, *38*(1), 179–202.

Kirschenman, J., & Neckerman, K. M. (1991). We'd love to hire them, but...": The meaning of race for employers. In F. L. Pincus & H. J. Ehrlich (Eds.), *Race and Ethnic Conflicts: Contending Views on Prejudice, Discrimination, and Ethnoviolence* (1st ed.). New York, NY: Routledge.

Kleist, J. S. (2001). *The effect of employment gaps on post-gap employability and earnings* (Publication No. 3031988) [Doctoral Dissertation, University of Minnesota]. ProQuest Dissertations Publishing.

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology, 100*(1), 128–161.

Koopmans, R., Veit, S., & Yemane, R. (2019). Taste or statistics? A correspondence study of ethnic, racial and religious labour market discrimination in Germany. *Ethnic and Racial Studies, 42*(16), 233–252.

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

Kuhfeld, W. F. (1997). Efficient experimental designs using computerized searches. In Sawtooth Software (Ed.). *Proceedings in the Sawtooth Software Conference August* 1997. Sequim, WA: Sawtooth Software.

Kuhfeld, W. F. (2010). *Statistical graphics in SAS: An introduction to the graph template language and the statistical graphics procedures*. Cary, NC: SAS Publishing.

Kuhfeld, W. F., Tobias, R. D., & Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research, 31*(4), 545–557.

Lahey, J. N. (2008). Age, women, and hiring an experimental study. *Journal of Human Resources, 43*(1), 30–56.

Landrine, H. (1985). Race x class stereotypes of women. *Sex Roles, 13*(1/2), 65–75.

Levinson, R. M. (1975). Sex discrimination and employment practices: An experiment with unconventional job inquiries. *Social Problems, 22*(4), 533–543.

Liebe, U., Moumouni, I. M., Bigler, C., Ingabire, C., & Bieri, S. (2020). Using factorial survey experiments to measure attitudes, social norms, and fairness concerns in developing countries. *Sociological methods & research, 49*(1), 161–192

Lippens, L., Vermeiren, S., & Baert, S. (2021). *The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments* (GLO Discussion Paper No. 972*).* Global Labor Organization.

Lucas, B. J., Berry, Z., Giurge, L. M., & Chugh, D. (2021). A longer shortlist increases the consideration of female candidates in male-dominant domains. *Nature Human Behaviour*, *5,* 736–742.

Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology, 91*(4), 777–785.

Macrae, C. N & Quadflieg, S. (2010). Perceiving people. In S. T. Fiske, D. T. Gilbert, & G. Lindzey, *Handbook of social psychology* (5th ed.). New York, NY: John Wiley.

Marreiros, H., Tonin, M., Vlassopoulos, M., & Schraefel, M. C. (2017). "Now that you mention it": A survey experiment on information, inattention and online privacy. *Journal of Economic Behavior & Organization*, 140, 1–17.

Mathew, S. (2017). How the second-order free rider problem is solved in a small-scale society. *American Economic Review, 107*(5), 578–581.

Mavlikeeva, M., & Asanov, I. (2020). Can Group Identity Explain Gender Gap in Recruitment Process? *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2020: Gender Economics, ZBW*. Leibniz Information Centre for Economics.

McDonald, P. (2019). How factorial survey analysis improves our understanding of employer preferences. *Swiss Journal of Sociology*, *45*(2), 237–260.

Merriweather, T. J. (2020). *A [K] ink in the Armor: How the Intersection of Gender and Racial Prototypicality Affect Perceptions of Black Women Aspiring to be Managers* [Doctoral dissertation, Columbia University]. ProQuest Dissertations Publishing.

Morgan, W. B., Walker, S. S., Hebl, M. M. R., & King, E. B. (2013). A field experiment: Reducing interpersonal discrimination toward pregnant job applicants. *Journal of Applied Psychology, 98*(5), 799.

Moshavi, D. (2004). He Said, She Said: Gender Bias and Customer Satisfaction With Phone-Based Service Encounters. *Journal of Applied Social Psychology, 34*(1), 162–176.

Namingit, S., Blankenau, W., & Schwab, B. (2021). Sick and tell: A field experiment analyzing the effects of an illness-related employment gap on the callback rate. *Journal of Economic Behavior & Organization, 185*, 865–882.

National Center for O*NET Development. O*NET OnLine. Retrieved from https://www.onetonline.org/.

Niemann, Y. F., Jennings, L., Rozelle, R. M., Baxter, J. C., & Sullivan, E. (1994). Use of free responses and cluster analysis to determine stereotypes of eight groups. *Personality and Social Psychology Bulletin, 20*(4), 379–390.

Nisic, N., & Auspurg, K. (2009). Faktorieller Survey und klassische Bevölkerungsumfrage im Vergleich—Validität, Grenzen und Möglichkeiten beider Ansätze. In P. Kriwy, C. Gross, & M. Jungbauer-Gans (Eds.), *Klein aber fein! Quantitative empirische Sozialforschung mit kleinen Fallzahlen* (pp. 211-245). VS Verlag für Sozialwissenschaften.

Nuijten, M. P., Poell, R. F., & Alfes, K. (2017). Extracurricular activities of Dutch university students and their effect on employment opportunities as perceived by both students and organizations. *International Journal of Selection and Assessment, 25*(4), 360–370.

Olian, J. D., Schwab, D. P., & Haberfeld, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of experimental studies. *Organizational Behavior and Human Decision Processes, 41*(2), 180–195.

Oreopoulos, P. (2011). Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand résumés. *American Economic Journal: Economic Policy, 3*(4), 148–71.

Pager, D., & Quillian, L. (2005). Walking the talk? What employers say versus what they do. *American Sociological Review, 70*(3), 355–380.

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163.

Peer, E., Rothschild, D. M., Evernden, Z., Gordon, A., & Damer, E. (2021). *MTurk, Prolific or panels? Choosing the right audience for online research. Choosing the right audience for online research* (SSRN Working Paper No. 103758). Social Science Research Network.

Peitz, L., Dhont, K., & Seyd, B. (2018). The psychology of supranationalism: Its ideological correlates and implications for EU attitudes and post-Brexit preferences. *Political Psychology, 39*(6), 1305–1322.

Petit, P. (2007). The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector. *Labour Economics, 14*(3), 371–391.

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

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of women quarterly, 24*(1), 81–92.

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex roles*, *59*(5), 377–391.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*(1), 119–125.

Riach, P., & Rich, J. (2004). Fishing for discrimination. *Review of Social Economy*, *62*(4), 465–486.

Riach, P. A., & Rich, J. (2006). An experimental investigation of sexual discrimination in hiring in the English labor market. *Advances in Economic Analysis & Policy, 5*(2), 1–20.

Ridgeway, C. L., & Kricheli-Katz, T. (2013). Intersecting cultural beliefs in social relations: Gender, race, and class binds and freedoms. *Gender & Society, 27*(3), 294–318.

Rivera, L. A., & Tilcsik, A. (2016). Class advantage, commitment penalty: The gendered effect of social class signals in an elite labor market. *American Sociological Review, 81*(6), 1097–1131.

Rosette, A. S., Ponce de Leon, R. , Koval, C. Z., & Harrison, D. A. (2018). Intersectionality: Connecting experiences of gender with race at work. *Research in Organizational Behavior*, 38, 1–22.

Rossi, P. H. (1951). *The Application of Latent Structure Analysis to the Study of Social Stratification*

[Unpublished doctoral Dissertation]. Columbia University.

Rossi, P. H., & Nock, S. L. (1982). *Measuring social judgements: The factorial survey approach.* Thousand Oaks, CA: Sage.

Roster, C. A., & Ferrari, J. R. (2020). Does work stress lead to office clutter, and how? Mediating influences of emotional exhaustion and indecision. *Environment and Behavior, 52*(9), 923–944.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of personality and social psychology, 77*(5), 1004–1010.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165–179.

Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., & Matthew Sobek, M. (2021). *IPUMS USA: Version 11.0 [American Community Survey]*. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D010.V11.0.

Sârbescu, P., Costea, I., & Rusu, S. (2011). Psychometric properties of the Marlowe-Crowne Social Desirability Scale in a Romanian sample. *Procedia—Social and Behavioral Sciences, 33*(1), 707–711

Sauer, C., Auspurg, K., & Hinz, T. (2020). Designing multi-factorial survey experiments: Effects of presentation style (text or table), answering scales, and vignette order. *Methods, Data, Analyses, 14*(2), 195–214.

Sauer, C., Auspurg, K., Hinz, T., & Liebig, S. (2011). The application of factorial surveys in general population samples: The effects of respondent age and education on response times and response consistency. *Survey Research Methods, 5*(3), 89–102.

Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of applied psychology, 57*(2), 95–100.

Schein, V. E. (1975). Relationships between sex role stereotypes and requisite management characteristics among female managers. *Journal of applied psychology, 60*(3), 340.

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology*, *46*(2), 356–360.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, NY: Houghton Mifflin Company.

Shahani-Denning, C., Andreoli, N., Snyder, J., Tevet, R., & Fox, S. (2011). The effects of physical attractiveness and gender on selection decisions: An experimental study. *International Journal of Management*, *28*(4), 16–23.

Shamon, H., Dülmer, H., & Giza, A. (2022). The factorial survey: the impact of the presentation format of vignettes on answer behavior and processing time. *Sociological Methods & Research*, *51*(1), 396–

438.

Sharpe Wessling, K., Huber, J., & Netzer, O. (2017). MTurk character misrepresentation: Assessment and solutions. *Journal of Consumer Research*, *44*(1), 211–230.

Singh, J., Crisafulli, B., Quamina, L. T., & Kottasz, R. (2020). The role of brand equity and crisis type on corporate brand alliances in crises. *European Management Review, 17*(4), 821–834.

Social Security Administration (2021). Popular Baby Names. Retrieved from: https://www.ssa.gov/OACT/babynames/

Sterkens, P., Baert, S., Rooman, C., & Derous, E. (2021). As if it weren't hard enough already: Breaking down hiring discrimination following burnout. Economics & Human Biology, 43, 101050.

Tajfel, H., & Turner, J. C. (1985). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (2nd ed., pp 7-24). Chicago, IL: Nelson-Hall.

Tosi, H. L., & Einbender, S. W. (1985). The effects of the type and amount of information in sex discrimination research: A meta-analysis. *Academy of Management Journal*, *28*(3), 712–723.

Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, *104*(2), 207–223.

U.S. Bureau of Labor Statistics (2021a). 100 years of working women. Retrieved from: https://www.dol.gov/agencies/wb/data/occupations-decades-100.

U.S. Bureau of Labor Statistics (2021b). Women in the labor force. Retrieved from: https://www.dol.gov/agencies/wb/data/facts-over-time/women-in-the-labor-force.

U.S. Census Bureau (2021a). Frequently Occurring Surnames from the 2010 Census. Retrieved from: https://www.census.gov/topics/population/genealogy/data/2010_surnames.html

U.S. Census Bureau (2021b). Quick Facts. Retrieved from: https://www.census.gov/quickfacts/fact/table/US/PST045219.

Valfort, M. A. (2020). Anti-Muslim discrimination in France: Evidence from a field experiment. *World Development, 135*, 105022.

Van Belle, E., Caers, R., Cuypers, L., De Couck, M., Neyt, B., Van Borm, H., & Baert, S. (2020). What do student jobs on graduate CVs signal to employers? *Economics of Education Review, 75*, 101979.

Van Belle, E., Caers, R., De Couck, M., Di Stasio, V., & Baert, S. (2019). The signal of applying for a job under a vacancy referral scheme. *Industrial Relations, 58*(2), 251–274.

Van Belle, E., Di Stasio, V., Caers, R., De Couck, M., & Baert, S. (2018). Why are employers put off by long spells of unemployment? *European Sociological Review, 34*(6), 694–710.

Van Borm, H. & Baert, S. (2021). *An Arab, An Asian, and A Black Guy Walk into a Job Interview: Examining Explanations for Ethnic Discrimination in Hiring*. Unpublished manuscript. Ghent University.

Van Borm, H., Burn, I., & Baert, S. (2021). What Does a Job Candidate's Age Signal to Employers? *Labour Economics*, 102003.

Van Borm, H., Dhoop, M., Van Acker, A., & Baert, S. (2020). What does someone's gender identity signal to employers? *International Journal of Manpower, 41*(6), 753–777.

Van Hoye, G., & Lievens, F. (2003). The effects of sexual orientation or hirability ratings: An experimental study. *Journal of Business and Psychology, 18*(1), 15–30.

Van Ommeren, J., Russo, G., De Vries, R. E., & Van Ommeren, M. (2005). Context in selection of men and women in hiring decisions: Gender composition of the applicant pool. *Psychological reports, 96*(2), 349–360.

Veit, S., & Thijsen, L. (2021). Almost identical but still treated differently: hiring discrimination against foreign-born and domestic-born minorities. *Journal of Ethnic and Migration Studies, 47*(6), 1285–1304.

Visser, I. (2002). Prototypes of gender: conceptions of feminine and masculine. *Women's studies international forum*, *25*(5), 529-539.

Wallace, M., Wright, B. R., & Hyde, A. (2014). Religious affiliation and hiring discrimination in the American South: A field experiment. *Social Currents, 1*(2), 189–207.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Science Research, 38*(3), 505–520.

Webster, S. W., O'Toole, R., O'Toole, A. W., & Lucal, B. (2005). Overreporting and underreporting of child abuse: Teachers' use of professional discretion. *Child abuse & neglect, 29*(11), 1281–1296.

Weichselbaumer, D. (2004). Is it sex or personality? The impact of sex stereotypes on discrimination in applicant selection. *Eastern Economic Journal, 30*(2), 159–186.

Weichselbaumer, D. (2020). Multiple discrimination against female immigrants wearing headscarves. *ILR Review*, *73*(3), 600-627.

Weisshaar, K. (2018). From opt out to blocked out: The challenges for labor market re-entry after family-related employment lapses. *American Sociological Review, 83*(1), 34–60.

Weitz, R., & Gordon, L. (1993). Images of Black women among Anglo college students. *Sex Roles, 28*(1), 19–34.

West, C. M. (2008). Mammy, Jezebel, Sapphire, and their homegirls: Developing an ''oppositional gaze'' toward the images of Black women. In J. C. Chrisler, C. Golden, & P. D. Rozee (Eds.), *Lectures on the psychology of women* (pp. 287–299). Boston, MA: McGraw-Hill

White, D. G. (1985). *Ar'n't I a woman: Female slaves in the plantation South*. New York, NY: Norton and Company.

Yemane, R. (2020). Cumulative disadvantage? The role of race compared to ethnicity, religion, and non-white phenotype in explaining hiring discrimination in the US labour market. *Research in Social Stratification and Mobility, 69*, 100552.

Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of* *personality and social psychology, 112*(5), 696.

Appendix

<Table A–1 about here>

<Table A–2 about here>

<Table A–3 about here>

<Table A–4 about here>

Table 1. Vignette factors and corresponding levels used in the experimental materials.

Vignette factors	Vignette levels
Name	{2× Name of African America man, 2× Name of African American woman, 4× Name of White Anglo-Saxon man, 4× Name of White Anglo-Saxon woman}
Commuting distance	{0–5 miles, 5–10 miles, 10–50 miles, More than 50 miles}
Experience in the occupation	{None, About 2 years, About 5 years, About 10 years}
Recent employment gap	{None, Due to unemployment, Due to health issues, Due to family responsibilities}
Extracurricular activities	{None, Volunteering, Sport activities, Cultural activities}

Note: The factorial product of the vignette levels (i.e., 12 × 4 × 4 × 4 × 4) resulted in 3,072 possible combinations. 80 sets of five vignettes were drawn from this vignette universe using a D-efficient Resolution V design (D-efficiency: 92.85; Auspurg & Hinz, 2014) and distributed at random to the participants, as described in Subsection 2.1.

Table 2. Statements used in the experimental materials.

Signals and evaluation outcome	Statements
Perceived intellectual abilities	'People with this profile usually have enough intellectual capacities to perform this job well.'
Perceived social abilities	'People with this profile usually have enough social capacities to perform this job well.'
Perceived physical abilities	'People with this profile usually have enough physical capacities to perform this job well.'
Perceived assertiveness	'People with this profile usually are assertive enough to perform this job well.'
Perceived ambition	'People with this profile usually are ambitious enough to perform this job well.'
Perceived motivation	'People with this profile usually are motivated enough to perform this job well.'
Perceived autonomy	'People with this profile usually are autonomous enough to perform this job well.'
Perceived resilience	'People with this profile usually are resilient enough to perform this job well.'
Perceived supportiveness	'People with this profile usually are supportive enough to perform this job well.'
Perceived reliability	'People with this profile usually are reliable enough to perform this job well.'
Perceived cooperativeness	'People with this profile usually are collaborative enough to perform this job well.'
Perceived flexibility	'People with this profile usually are flexible enough to perform this job well.'
Perceived openness to new people and experiences	'People with this profile usually are open enough to new people and experiences to perform this job well.'
Perceived respect towards authority	'People with this profile usually have enough respect towards authority to perform this job well.'
Perceived sense of responsibility	'People with this profile usually have a large enough sense of responsibility to perform this job well.'
Perceived degree of absence due to childcare responsibilities	'People with this profile usually fall out in the short or medium run due to childcare responsibilities'
Attitude towards collaboration of employer	'I think I usually would enjoy collaborating with people like this person.'
Attitude towards collaboration of other employees	'I think other employees usually would enjoy collaborating with people like this person.'
Attitude towards collaboration of customers	'I think customers usually would enjoy collaborating with people like this person.'
Invitation probability	'I advise to invite this candidate for the second phase of the application process.'

Note: In this table, we present the potential gender stereotypes and attitudes towards women, the invitation to interview probability, and their corresponding statements as they were included in the online experiment. The participants evaluated each statement on an 11-point Likert scale ranging from 0 (i.e., 'strongly disagree') to 10 (i.e., 'strongly agree').

Table 3. Data description by fictitious candidate's gender.

	Total sample [N = 1,449]	Candidate's gender: Male [N = 725]	Candidate's gender: Female [N = 724]	Difference (2) – (3)
	(1)	(2)	(3)	(4)
A. PARTICIPANT CHARACTERISTICS				
Gender: female	0.435	0.446	0.424	0.022 [0.834]
Age: < 41 years old	0.621	0.616	0.626	−0.009 [−0.358]
Race-ethnicity: white Anglo-Saxon	0.828	0.822	0.834	−0.012 [−0.614]
Region of residence: South	0.336	0.343	0.330	0.013 [0.537]
Highest educational degree: university	0.780	0.775	0.784	−0 .009 [−0.430]
Employment status: full time employment	0.810	0.801	0.819	−0.018 [−0.858]
Frequency of hiring: ≥ once per semester	0.697	0.683	0.711	−0.029 [−1.183]
Experience as HR professional: <= 1 year	0.917	0.923	0.912	0.011 [0.770]
B. JOB CHARACTERISTICS				
Ophthalmic laboratory technician	0.082	0.079	0.086	−0.007 [−0.486]
Welder	0.068	0.068	0.068	−0.000 [−0.422]
Payroll and timekeeping clerk	0.079	0.074	0.084	−0.010 [−0.688]
Production, planning, and expediting clerk	0.076	0.080	0.072	0 .008 [0.587]
Door-to-door sales worker	0.080	0.090	0.070	0 .019 [1.347]
Food-batch-maker	0.073	0.076	0.070	0.005 [0.396]
Operations research analyst	0.072	0.070	0.075	−0.004 [−0.311]
Chemical engineer	0.065	0.070	0.059	0.011 [0.846]
Proof reader	0.050	0.055	0.044	0.011 [0.961]
Human resources manager	0.074	0.069	0.079	−0.010 [−0.710]
Insurance sales agent	0.050	0.044	0.055	−0.011 [−0.973]
Wind farm support specialist	0.081	0.076	0 .087	−0.011 [−0.776]

Note: *T*-tests are performed to test whether the differences between the subsamples by candidate's gender are significantly different from 0. X^2-tests, which are more appropriate for binary outcomes, yield exactly the same conclusions. *** (**) ((*)) indicates significance at 1% (5%) ((10%)) significance level. *T*-statistics are in brackets. *N* = 1,441 (i.e., 814 men and 627 women) for the *T*-test related to participants' gender because two participants identified themselves as non-binary (i.e., 'X'). Observations with respect to these participants were therefore not included in the analysis.

Table 4. Moderation analysis with interview probability as outcome variable.

	(1)	(2)	(3)	(4)	(5)
A. Candidate characteristics					
Gender: female	0.007 (0.036)	0.159* (0.104)	−0.031 (0.126)	−0.039 (0.050)	0.071 (0.174)
Race: black	0.014 (0.043)	0.035 (0.057)	0.013 (0.043)	0.015 (0.043)	0.035 (0.057)
Commuting distance: more than 50 miles	−0.421*** (0.052)	−0.344*** (0.066)	−0.422*** (0.052)	−0.422*** (0.052)	−0.343*** (0.067)
Experience in the occupation: none	−1.205*** (0.053)	−1.165*** (0.073)	−1.205*** (0.053)	−1.198*** (0.054)	−1.161*** (0.074)
Recent employment gap					
Due to unemployment	−0.250*** (0.053)	−0.186* (0.080)	−0.249*** (0.053)	−0.258*** (0.054)	−0.193** (0.080)
Due to health issues	−0.306*** (0.055)	−0.315*** (0.081)	−0.305*** (0.055)	−0.306*** (0.055)	−0.316*** (0.082)
Due to family responsibilities	−0.221*** (0.052)	−0.206** (0.079)	−0.221*** (0.052)	−0.221*** (0.052)	−0.208** (0.079)
None (reference)	-	-	-	-	-
Extracurricular activities (ECA)					
Volunteering	0.119** (0.050)	0.161** (0.079)	0.120** (0.050)	0.115** (0.050)	0.159* (0.079)
Sport activities	0.081 (0.055)	0.127 (0.075)	0.081 (0.055)	0.080 (0.055)	0.128 (0.075)
Cultural activities	0.065 (0.049)	0.066 (0.075)	0.066 (0.049)	0.062 (0.050)	0.065 (0.075)
None (reference)	-	-	-	-	-
Candidate’s gender: Female x Race: black	-	−0.051 (0.080)	-	-	−0.053 (0.080)
Candidate’s gender: Female x Commuting distance: more than 50 miles	-	−0.154 (0.094)	-	-	−0.161 (0.094)
Candidate’s gender: Female x Experience: none	-	−0.083 (0.109)	-	-	−0.075 (0.110)
Candidate’s gender: Female x Employment gap: due to unemployment	-	−0.130 (0.113)	-	-	−0.132 (0.113)
Candidate’s gender: Female x Employment gap: due to health issues	-	0.025 (0.126)	-	-	0.027 (0.126)
Candidate’s gender: Female x Employment gap: due to family responsibilities	-	−0.032 (0.105)	-	-	−0.029 (0.105)
Candidate’s gender: Female x ECA: Volunteering	-	−0.078 (0.121)	-	-	−0.079 (0.121)
Candidate’s gender: Female x ECA: Sport activities	-	−0.093 (0.108)	-	-	−0.100 (0.108)
Candidate’s gender: Female x ECA: Cultural activities	-	0.003 (0.125)	-	-	0.001 (0.125)
Observations					1,449

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression models outlined in Subsection 4.1. More concretely, in column (1), we present the coefficient estimates of our baseline model. In column (2), (3), (4), and (5), we present the coefficient estimates of four models in which we, subsequently, add different interaction terms to the baseline model. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table 4. Moderation analysis with interview probability as outcome variable (Continued).

	(1)	(2)	(3)	(4)	(5)
B. Applicant pool characteristics					
Gender composition applicant pool: majority is male (GCAP)	0.049 (0.054)	0.055 (0.054)	0.049 (0.054)	0.001 (0.063)	0.002 (0.063)
Candidate's gender: Female x GCAP: majority is male	-	-	-	0.146* (0.077)	0.161** (0.078)
C. Job characteristics					
Level of required skills in occupation: high	0.006 (0.010)	0.006 (0.010)	0.007 (0.013)	0.006 (0.010)	0.008 (0.014)
Level of required customer contact in occupation: high	0.003 (0.008)	0.003 (0.008)	0.001 (0.009)	0.003 (0.008)	0.001 (0.009)
Level of required physical effort in occupation: high	0.020* (0.009)	0.019* (0.009)	0.020* (0.011)	0.020* (0.009)	0.020* (0.011)
Level of required leadership skills in occupation: high	−0.013 (0.012)	−0.013 (0.012)	−0.022 (0.014)	−0.013 (0.012)	−0.022 (0.014)
Gender composition in occupation: male	−0.018 (0.013)	−0.018 (0.014)	−0.015 (0.015)	−0.018 (0.013)	−0.015 (0.016)
Candidate's gender: Female x Level of required skills: high	-	-	−0.001 (0.015)	-	−0.003 (0.015)
Candidate's gender: Female x Level of required customer contact: high	-	-	0.003 (0.011)	-	0.004 (0.011)
Candidate's gender: Female x Level of required physical effort: high	-	-	−0.001 (0.013)	-	−0.002 (0.013)
Candidate's gender: Female x Level of required leadership skills: high	-	-	0.017 (0.016)	-	0.018 (0.016)
Candidate's gender: Female x Gender composition in occupation: male	-	-	−0.007 (0.018)	-	−0.006 (0.018)
Observations					1,449

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression models outlined in Subsection 4.1. More concretely, in column (1), we present the coefficient estimates of our baseline model. In column (2), (3), (4), and (5), we present the coefficient estimates of four models in which we, subsequently, add different interaction terms to the baseline model. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table 5. Multivariate regression analysis with stereotype and attitude scales as outcome variable.

	Intellectual abilities	Social abilities	Physical abilities	Assertiveness	Ambition	Motivation	Autonomy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Candidate characteristics							
Gender: female	0.053 (0.038)	0.090** (0.038)	−0.129*** (0.043)	−0.077** (0.041)	0.046 (0.040)	0.046 (0.039)	−0.032 (0.042)
Race: black	−0.003 (0.046)	0.065* (0.048)	0.036 (0.045)	0.055 (0.048)	0.034 (0.047)	0.081** (0.046)	0.057 (0.044)
Commuting distance: more than 50 miles (CD)	−0.097** (0.047)	−0.110** (0.046)	−0.094** (0.046)	−0.049 (0.044)	−0.102** (0.048)	−0.131*** (0.047)	−0.072* (0.047)
Experience in the occupation: none	−0.849*** (0.056)	−0.570*** (0.055)	−0.485*** (0.053)	−0.721*** (0.057)	−0.716*** (0.057)	−0.758*** (0.056)	−0.92*** (0.061)
Recent employment gap (EG)							
Due to unemployment	−0.182*** (0.050)	−0.174** (0.054)	−0.136** (0.048)	−0.175** (0.058)	−0.257*** (0.056)	−0.289*** (0.055)	−0.178** (0.053)
Due to health issues	−0.172*** (0.051)	−0.149*** (0.053)	−0.864*** (0.062)	−0.185*** (0.055)	−0.248*** (0.050)	−0.262*** (0.053)	−0.206*** (0.056)
Due to family responsibilities	−0.145** (0.052)	−0.117 (0.055)	−0.163*** (0.046)	−0.114* (0.052)	−0.252*** (0.057)	−0.225*** (0.056)	−0.139** (0.053)
None (reference)	-	-	-	-	-	-	-
Extracurricular activities (ECA)							
Volunteering	0.181*** (0.052)	0.379*** (0.057)	0.158** (0.054)	0.170*** (0.053)	0.221*** (0.057)	0.218*** (0.055)	0.064 (0.056)
Sport activities	0.123* (0.050)	0.325*** (0.057)	0.394*** (0.056)	0.240*** (0.056)	0.254*** (0.057)	0.240*** (0.055)	0.153** (0.056)
Cultural activities	0.139** (0.053)	0.278*** (0.057)	0.136 (0.055)	0.134* (0.054)	0.161** (0.056)	0.154** (0.056)	0.071 (0.056)
None (reference)	-	-	-	-	-	-	-
B. Applicant pool characteristics							
Gender composition applicant pool: majority is male (GCAP)	0.229 (0.137)	0.036 (0.075)	−0.052 (0.066)	−0.010 (0.073)	0.021 (0.072)	0.009 (0.070)	−0.025 (0.068)
C. Job characteristics							
Level of required skills in occupation: high (RS)	−0.014 (0.015)	0.001 (0.015)	−0.016 (0.013)	0.003 (0.015)	0.006 (0.014)	−0.000 (0.014)	−0.008 (0.014)
Level of required customer contact in occupation: high (RCC)	−0.018 (0.012)	−0.018 (0.012)	−0.020** (0.010)	−0.008 (0.012)	−0.003 (0.011)	−0.008 (0.012)	−0.011 (0.011)
Level of required physical effort in occupation: high (RPE)	−0.012 (0.013)	−0.002 (0.013)	−0.015 (0.011)	−0.005 (0.013)	−0.006 (0.013)	−0.007 (0.013)	0.006 (0.012)
Level of required leadership skills in occupation: high (RLS)	0.004 (0.016)	0.008 (0.017)	0.000 (0.015)	0.018 (0.017)	0.007 (0.016)	0.009 (0.016)	0.003 (0.015)
Gender composition in occupation: male (GCO)	0.028 (0.021)	0.028 (0.022)	−0.006 (0.018)	0.013 (0.020)	0.008 (0.020)	0.012 (0.020)	0.010 (0.019)
Observations							1,449

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression models outlined in Subsection 4.2. More concretely, in each column, we present the coefficient estimates of the regression where we regress one of the stereotypes or attitude scales on (i) all candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table 5. Multivariate regression analysis with stereotype and attitude scales as outcome variable (Continued).

	Resilience	Supportiveness	Reliability	Cooperativeness	Flexibility	Creativity	Accuracy
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
A. CANDIDATE CHARACTERISTICS							
Gender: female	−0.036 (0.038)	0.110*** (0.038)	0.032 (0.041)	0.052 (0.037)	−0.013 (0.041)	0.090** (0.038)	0.073 (0.037)
Race: black	0.039 (0.045)	0.010 (0.047)	0.046 (0.048)	0.030 (0.045)	0.002 (0.049)	0.034 (0.044)	0.011 (0.045)
Commuting distance: more than 50 miles (CD)	−0.162*** (0.050)	−0.098** (0.048)	−0.302*** (0.052)	−0.114*** (0.046)	−0.298*** (0.051)	−0.079** (0.043)	−0.064** (0.044)
Experience in the occupation: none	−0.696*** (0.056)	−0.688*** (0.055)	−0.741*** (0.052)	−0.672*** (0.054)	−0.618*** (0.052)	−0.707*** (0.054)	−0.981*** (0.057)
Recent employment gap (EG)							
Due to unemployment	−0.256*** (0.054)	−0.157** (0.054)	−0.300*** (0.055)	−0.178*** (0.051)	−0.199*** (0.052)	−0.175** (0.053)	−0.217*** (0.051)
Due to health issues	−0.445*** (0.057)	−0.135*** (0.051)	−0.513*** (0.057)	−0.154*** (0.052)	−0.350*** (0.054)	−0.135** (0.053)	−0.217*** (0.050)
Due to family responsibilities	−0.252*** (0.055)	−0.063 (0.054)	−0.469*** (0.06)	−0.131** (0.053)	−0.477*** (0.061)	−0.107 (0.050)	−0.123* (0.051)
None (reference)	-	-	-	-	-	-	-
Extracurricular activities (ECA)							
Volunteering	0.234*** (0.056)	0.246*** (0.054)	0.202*** (0.055)	0.279*** (0.053)	0.199*** (0.056)	0.200*** (0.052)	0.151** (0.052)
Sport activities	0.298*** (0.057)	0.151** (0.057)	0.194** (0.056)	0.253*** (0.055)	0.202*** (0.059)	0.154** (0.052)	0.130* (0.051)
Cultural activities	0.149* (0.058)	0.101 (0.055)	0.134* (0.054)	0.182*** (0.053)	0.129* (0.059)	0.220*** (0.053)	0.054 (0.054)
None (reference)	-	-	-	-	-	-	-
B. APPLICANT POOL CHARACTERISTICS							
Gender composition applicant pool: majority is male (GCAP)	−0.009 (0.067)	−0.007 (0.075)	−0.038 (0.063)	−0.005 (0.074)	−0.023 (0.068)	−0.008 (0.077)	−0.031 (0.068)
C. JOB CHARACTERISTICS							
Level of required skills in occupation: high (RS)	0.004 (0.013)	0.006 (0.016)	0.017 (0.013)	0.006 (0.015)	0.016 (0.013)	0.017 (0.016)	−0.009 (0.014)
Level of required customer contact in occupation: high (RCC)	−0.014 (0.010)	−0.019* (0.012)	−0.001 (0.010)	−0.014 (0.012)	−0.006 (0.011)	−0.004 (0.013)	−0.013 (0.011)
Level of required physical effort in occupation: high (RPE)	0.004 (0.012)	−0.000 (0.013)	0.012 (0.011)	−0.006 (0.014)	−0.005 (0.013)	0.012 (0.014)	−0.007 (0.012)
Level of required leadership skills in occupation: high (RLS)	0.017 (0.014)	0.011 (0.017)	0.004 (0.013)	0.015 (0.017)	0.002 (0.015)	0.001 (0.018)	0.001 (0.015)
Gender composition in occupation: male (GCO)	0.010 (0.019)	0.024 (0.021)	−0.001 (0.017)	0.033 (0.022)	0.009 (0.019)	0.005 (0.022)	0.008 (0.019)
Observations							1,449

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression models outlined in Subsection 4.2. More concretely, in each column, we present the coefficient estimates of the regression where we regress one of the stereotypes or attitude scales on (i) all candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table 5. Multivariate regression analysis with stereotype and attitude scales as outcome variable (Continued).

	Openness	Respect towards authority	Sense of responsibility	Absence due to family responsibilities	Willingness to collaborate: employers	Willingness to collaborate: co-workers	Willingness to collaborate: Customers
	(15)	(16)	(17)	(18)	(19)	(20)	(21)
A. Candidate characteristics							
Gender: female	0.104** (0.041)	0.027 (0.038)	0.060 (0.038)	0.255*** (0.046)	0.102*** (0.039)	0.076** (0.038)	0.091** (0.037)
Race: black	0.045 (0.049)	−0.046 (0.049)	0.022 (0.048)	−0.029 (0.045)	0.018 (0.046)	0.018 (0.046)	0.058 (0.048)
Commuting distance: more than 50 miles (CD)	−0.086** (0.047)	−0.009 (0.045)	−0.074** (0.046)	0.081 (0.045)	−0.109*** (0.047)	−0.133*** (0.048)	−0.105** (0.048)
Experience in the occupation: none	−0.505*** (0.057)	−0.553*** (0.050)	−0.777*** (0.052)	−0.046 (0.046)	−0.686*** (0.053)	−0.699*** (0.053)	−0.610*** (0.053)
Recent employment gap (EG)							
Due to unemployment	−0.170*** (0.057)	−0.171*** (0.056)	−0.287*** (0.054)	0.053 (0.052)	−0.185*** (0.053)	−0.132* (0.054)	−0.179*** (0.055)
Due to health issues	−0.161*** (0.056)	−0.053 (0.052)	−0.220*** (0.051)	0.139** (0.052)	−0.166*** (0.051)	−0.106** (0.053)	−0.168*** (0.055)
Due to family responsibilities	−0.147** (0.054)	−0.066 (0.050)	−0.171** (0.056)	0.918*** (0.069)	−0.101 (0.055)	−0.106 (0.054)	−0.088 (0.055)
None (reference)	-	-	-	-	-	-	-
Extracurricular activities (ECA)							
Volunteering	0.291*** (0.057)	0.167*** (0.053)	0.213*** (0.055)	−0.073 (0.059)	0.261*** (0.055)	0.365*** (0.054)	0.327*** (0.056)
Sport activities	0.216*** (0.056)	0.150** (0.052)	0.205*** (0.056)	−0.047 (0.055)	0.238*** (0.055)	0.346*** (0.058)	0.28*** (0.056)
Cultural activities	0.253*** (0.058)	0.064 (0.053)	0.121 (0.055)	−0.023 (0.056)	0.148** (0.056)	0.261*** (0.055)	0.207*** (0.055)
None (reference)	-	-	-	-	-	-	-
B. Applicant pool characteristics							
Gender composition applicant pool: majority is male (GCAP)	−0.003 (0.075)	−0.035 (0.081)	−0.001 (0.068)	−0.049 (0.066)	0.026 (0.075)	0.024 (0.075)	0.012 (0.077)
C. Job characteristics							
Level of required skills in occupation: high (RS)	0.005 (0.015)	0.007 (0.017)	0.003 (0.014)	0.021* (0.012)	0.014 (0.016)	0.007 (0.015)	0.011 (0.016)
Level of required customer contact in occupation: high (RCC)	−0.015 (0.012)	−0.021 (0.013)	−0.015 (0.011)	0.018* (0.010)	−0.009 (0.012)	−0.017 (0.012)	−0.010 (0.012)
Level of required physical effort in occupation: high (RPE)	0.003 (0.013)	0.009 (0.014)	0.004 (0.012)	0.041*** (0.011)	0.002 (0.014)	0.008 (0.013)	0.007 (0.013)
Level of required leadership skills in occupation: high (RLS)	0.014 (0.017)	0.008 (0.018)	0.003 (0.015)	−0.000 (0.014)	−0.007 (0.017)	0.011 (0.017)	0.011 (0.018)
Gender composition in occupation: male (GCO)	0.016 (0.021)	0.018 (0.023)	0.016 (0.019)	−0.011 (0.014)	0.017 (0.021)	0.012 (0.020)	0.019 (0.021)
Observations							1,449

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression models outlined in Subsection 4.2. More concretely, in each column, we present the coefficient estimates of the regression where we regress one of the stereotypes or attitude scales on (i) all candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table 6. Moderation analysis with stereotype and attitude scales as outcome variable.

	Intellectual abilities	Social abilities	Physical abilities	Assertiveness	Ambition	Motivation	Autonomy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. CANDIDATE CHARACTERISTICS							
Gender: female	0.109 (0.190)	0.296 (0.194)	−0.109 (0.190)	0.125 (0.194)	−0.060 (0.202)	0.150 (0.193)	0.032 (0.188)
Race: black	0.068 (0.063)	0.133** (0.061)	0.083* (0.060)	0.096** (0.062)	0.066 (0.062)	0.115** (0.063)	0.094* (0.063)
Commuting distance: more than 50 miles (CD)	−0.037 (0.071)	−0.076 (0.068)	−0.076* (0.064)	−0.005 (0.067)	−0.071 (0.071)	−0.104* (0.070)	−0.040 (0.066)
Experience in the occupation: none	−0.848*** (0.078)	−0.554*** (0.078)	−0.450*** (0.070)	−0.720*** (0.081)	−0.732*** (0.075)	−0.767*** (0.078)	−0.879*** (0.082)
Recent employment gap (EG)							
Due to unemployment	−0.274** (0.081)	−0.368*** (0.085)	−0.228*** (0.068)	−0.219** (0.081)	−0.346*** (0.081)	−0.325*** (0.083)	−0.253** (0.081)
Due to health issues	−0.121 (0.081)	−0.201** (0.079)	−0.898*** (0.091)	−0.189** (0.083)	−0.263*** (0.079)	−0.207** (0.081)	−0.236*** (0.085)
Due to family responsibilities	−0.057 (0.082)	0.050 (0.080)	−0.069 (0.070)	0.005 (0.083)	−0.207* (0.083)	−0.114 (0.081)	−0.123 (0.078)
None (reference)	-	-	-	-	-	-	-
Extracurricular activities (ECA)							
Volunteering	0.185* (0.086)	0.459*** (0.093)	0.115 (0.086)	0.185** (0.089)	0.228*** (0.086)	0.237*** (0.089)	0.106 (0.085)
Sport activities	0.089 (0.077)	0.395*** (0.084)	0.357*** (0.077)	0.264*** (0.080)	0.264*** (0.080)	0.234** (0.079)	0.146 (0.077)
Cultural activities	0.122 (0.090)	0.356*** (0.093)	0.148 (0.090)	0.167 (0.088)	0.161 (0.093)	0.145 (0.092)	−0.031 (0.088)
None (reference)	-	-	-	-	-	-	-
Candidate’s gender: Female x Race: black	−0.154** (0.086)	−0.141* (0.083)	−0.100 (0.088)	−0.094 (0.084)	−0.074 (0.084)	−0.085 (0.081)	−0.090 (0.091)
Candidate’s gender: Female x CD: more than 50 miles	−0.128 (0.105)	−0.063 (0.101)	−0.045 (0.098)	−0.087 (0.102)	−0.066 (0.107)	−0.057 (0.107)	−0.065 (0.097)
Candidate’s gender: Female x Experience: none	0.011 (0.127)	0.034 (0.124)	−0.078 (0.109)	0.027 (0.118)	0.040 (0.116)	0.022 (0.119)	−0.084 (0.119)
Candidate’s gender: Female x EG: due to unemployment	0.178 (0.126)	0.396*** (0.131)	0.192 (0.106)	0.118 (0.123)	0.174 (0.137)	0.078 (0.133)	0.154 (0.121)
Candidate’s gender: Female x EG: due to health issues	−0.104 (0.123)	0.111 (0.122)	0.074 (0.125)	0.017 (0.125)	−0.032 (0.123)	−0.115 (0.124)	0.059 (0.128)
Candidate’s gender: Female x EG: due to family responsibilities	−0.191** (0.133)	−0.133* (0.124)	−0.191*** (0.117)	−0.223** (0.130)	−0.096 (0.133)	−0.232*** (0.121)	−0.039 (0.120)
Candidate’s gender: Female x ECA: Volunteering	0.006 (0.139)	−0.161 (0.144)	0.091 (0.131)	−0.017 (0.137)	−0.012 (0.133)	−0.024 (0.138)	−0.069 (0.129)
Candidate’s gender: Female x ECA: Sport activities	0.084 (0.131)	−0.140 (0.134)	0.080 (0.127)	−0.042 (0.131)	−0.017 (0.129)	0.026 (0.129)	0.019 (0.120)
Candidate’s gender: Female x ECA: Cultural activities	0.052 (0.137)	−0.147 (0.147)	0.002 (0.141)	−0.034 (0.147)	0.015 (0.147)	0.035 (0.143)	0.090 (0.137)
B. APPLICANT POOL CHARACTERISTICS							
Gender composition applicant pool: majority is male (GCAP)	−0.012 (0.086)	0.050 (0.084)	−0.045 (0.077)	0.059 (0.085)	−0.007 (0.084)	0.025 (0.085)	−0.015 (0.082)
Candidate’s gender: Female x GCAP: majority is male	−0.005 (0.079)	−0.079 (0.081)	−0.052 (0.088)	−0.167* (0.086)	0.039 (0.084)	−0.053 (0.082)	−0.048 (0.090)
C. JOB CHARACTERISTICS		4					
Level of required skills in occupation: high (RS)	−0.022 (0.017)	0.003 (0.017)	−0.023 (0.016)	−0.014 (0.017)	−0.006 (0.016)	0.002 (0.018)	−0.014 (0.018)

Level of required customer contact in occupation: high (RCC)	−0.021 (0.013)	−0.020 (0.014)	−0.029*** (0.013)	−0.011 (0.013)	−0.002 (0.012)	−0.007 (0.013)	−0.013 (0.013)
Level of required physical effort in occupation: high (RPE)	−0.017 (0.014)	−0.003 (0.014)	−0.017 (0.013)	−0.011 (0.015)	−0.010 (0.014)	−0.011 (0.014)	−0.005 (0.013)
Level of required leadership skills in occupation: high (RLS)	0.003 (0.019)	0.015 (0.019)	−0.012 (0.017)	0.021 (0.020)	0.002 (0.018)	0.004 (0.019)	0.000 (0.019)
Gender composition in occupation: male (GCO)	0.042* (0.022)	0.037 (0.024)	0.029 (0.021)	0.035 (0.022)	0.018 (0.022)	0.020 (0.022)	0.030 (0.020)
Candidate’s gender: Female x RS: high	0.017 (0.017)	0.008 (0.017)	0.017 (0.019)	0.035** (0.017)	0.024 (0.018)	−0.006 (0.017)	0.014 (0.018)
Candidate’s gender: Female x RCC: high	0.005 (0.014)	0.003 (0.015)	0.017 (0.014)	0.005 (0.013)	0.002 (0.013)	−0.001 (0.013)	0.002 (0.014)
Candidate’s gender: Female x RPE: high	0.013 (0.014)	0.004 (0.013)	0.004 (0.014)	0.012 (0.013)	0.008 (0.014)	0.010 (0.014)	0.022 (0.014)
Candidate’s gender: Female x RLS: high	0.006 (0.018)	−0.010 (0.019)	0.025 (0.019)	−0.002 (0.018)	0.014 (0.018)	0.010 (0.017)	0.008 (0.020)
Candidate’s gender: Female x GCO: male	−0.029 (0.021)	−0.018 (0.022)	−0.047* (0.023)	−0.045** (0.020)	−0.021 (0.020)	−0.018 (0.021)	−0.040 (0.022)
Observations							1,449

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression models outlined in Subsection 4.3. More concretely, in each column, we present the coefficient estimates of the regression where we regress one of the stereotypes or attitude scales on (i) all candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics, while including all interaction terms (i.e., interactions between the gender of the job applicant and (i) all other candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table 6. Moderation analysis with stereotype and attitude scales as outcome variable (Continued).

	Resilience	Supportiveness	Reliability	Cooperativeness	Flexibility	Creativity	Accuracy
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
A. CANDIDATE CHARACTERISTICS							
Gender: female	0.168 (0.187)	0.502*** (0.207)	0.144 (0.194)	0.183 (0.195)	0.139 (0.188)	0.067 (0.183)	0.228 (0.178)
Race: black	0.071 (0.060)	0.129** (0.058)	0.090 (0.064)	0.123** (0.057)	0.031 (0.063)	0.133** (0.057)	0.116*** (0.060)
Commuting distance: more than 50 miles (CD)	−0.062 (0.070)	−0.109* (0.069)	−0.279*** (0.071)	−0.061 (0.070)	−0.292*** (0.072)	−0.063 (0.067)	−0.056 (0.066)
Experience in the occupation: none	−0.671*** (0.075)	−0.602*** (0.076)	−0.744*** (0.070)	−0.656*** (0.073)	−0.647*** (0.073)	−0.712*** (0.076)	−1.001*** (0.078)
Recent employment gap (EG)							
Due to unemployment	−0.308*** (0.077)	−0.245** (0.084)	−0.398*** (0.081)	−0.301*** (0.080)	−0.230** (0.079)	−0.307*** (0.082)	−0.247** (0.081)
Due to health issues	−0.381*** (0.084)	−0.147* (0.079)	−0.472*** (0.078)	−0.174* (0.078)	−0.305*** (0.080)	−0.155* (0.082)	−0.211** (0.075)
Due to family responsibilities	−0.182 (0.079)	0.008 (0.080)	−0.408*** (0.084)	0.043 (0.080)	−0.359*** (0.083)	0.054 (0.074)	0.046 (0.080)
None (reference)	-	-	-	-	-	-	-
Extracurricular activities (ECA)							
Volunteering	0.293*** (0.088)	0.316*** (0.087)	0.204** (0.082)	0.303*** (0.085)	0.261*** (0.087)	0.210** (0.085)	0.134 (0.085)
Sport activities	0.338*** (0.078)	0.174 (0.082)	0.174* (0.076)	0.263*** (0.076)	0.234*** (0.081)	0.101 (0.077)	0.089 (0.076)
Cultural activities	0.127 (0.093)	0.160 (0.092)	0.120 (0.085)	0.150 (0.091)	0.084 (0.091)	0.204** (0.091)	0.023 (0.089)
None (reference)	-	-	-	-	-	-	-
Candidate's gender: Female x Race: black	−0.080 (0.087)	−0.237*** (0.086)	−0.093 (0.085)	−0.196** (0.080)	−0.082 (0.083)	−0.198** (0.086)	−0.223*** (0.080)
Candidate's gender: Female x CD: more than 50 miles	−0.195 (0.105)	0.021 (0.102)	−0.048 (0.103)	−0.108 (0.108)	0.014 (0.107)	−0.037 (0.104)	−0.026 (0.096)
Candidate's gender: Female x Experience: none	0.048 (0.119)	−0.170 (0.119)	0.003 (0.111)	−0.035 (0.118)	0.062 (0.116)	0.019 (0.123)	0.029 (0.121)
Candidate's gender: Female x EG: due to unemployment	0.105 (0.116)	0.170 (0.135)	0.201 (0.125)	0.250 (0.129)	0.068 (0.128)	0.253* (0.129)	0.078 (0.123)
Candidate's gender: Female x EG: due to health issues	−0.127 (0.123)	−0.030 (0.123)	−0.080 (0.119)	0.047 (0.124)	−0.093 (0.118)	−0.041 (0.124)	−0.012 (0.120)
Candidate's gender: Female x EG: due to family responsibilities	−0.150** (0.119)	−0.149** (0.127)	−0.126* (0.116)	−0.189** (0.125)	−0.246** (0.122)	−0.120 (0.121)	−0.170** (0.120)
Candidate's gender: Female x ECA: Volunteering	−0.096 (0.138)	−0.150 (0.141)	0.005 (0.130)	−0.039 (0.138)	−0.112 (0.137)	−0.019 (0.138)	0.044 (0.131)
Candidate's gender: Female x ECA: Sport activities	−0.067 (0.124)	−0.051 (0.135)	0.050 (0.124)	−0.009 (0.127)	−0.057 (0.131)	0.113 (0.124)	0.089 (0.121)
Candidate's gender: Female x ECA: Cultural activities	0.055 (0.142)	−0.117 (0.145)	0.043 (0.140)	0.085 (0.145)	0.101 (0.144)	0.048 (0.142)	0.090 (0.136)
B. APPLICANT POOL CHARACTERISTICS							
Gender composition applicant pool: majority is male (GCAP)	0.020 (0.079)	−0.005 (0.083)	−0.028 (0.078)	0.021 (0.082)	−0.010 (0.080)	−0.026 (0.086)	−0.013 (0.080)
Candidate's gender: Female x GCAP: majority is male	−0.050 (0.082)	0.017 (0.082)	−0.053 (0.087)	−0.074 (0.079)	−0.041 (0.086)	0.036 (0.081)	−0.111 (0.077)
C. JOB CHARACTERISTICS							
Level of required skills in occupation: high (RS)	−0.007 (0.016)	0.002 (0.017)	0.020 (0.015)	−0.008 (0.017)	0.017 (0.016)	0.007 (0.018)	−0.024 (0.017)
Level of required customer contact in occupation: high (RCC)	−0.018 (0.012)	−0.024* (0.013)	0.007 (0.012)	−0.018 (0.013)	−0.004 (0.013)	−0.001 (0.014)	−0.015 (0.012)

Level of required physical effort in occupation: high (RPE)	−0.006 (0.014)	−0.004 (0.014)	0.019 (0.012)	−0.012 (0.014)	−0.009 (0.013)	0.013 (0.015)	−0.015 (0.013)
Level of required leadership skills in occupation: high (RLS)	0.030* (0.017)	0.024 (0.019)	−0.014 (0.016)	0.029 (0.019)	−0.011 (0.018)	0.006 (0.020)	0.002 (0.018)
Gender composition in occupation: male (GCO)	0.022 (0.022)	0.042* (0.022)	−0.002 (0.019)	0.045* (0.023)	0.023 (0.021)	0.010 (0.023)	0.034 (0.020)
Candidate's gender: Female x RS: high	0.020 (0.017)	−0.008 (0.019)	−0.005 (0.017)	0.027 (0.017)	−0.000 (0.018)	0.020 (0.018)	0.029 (0.016)
Candidate's gender: Female x RCC: high	0.007 (0.013)	0.008 (0.013)	−0.014 (0.013)	0.006 (0.014)	−0.005 (0.015)	−0.007 (0.014)	0.002 (0.012)
Candidate's gender: Female x RPE: high	0.022 (0.013)	0.010 (0.014)	−0.013 (0.013)	0.013 (0.014)	0.008 (0.014)	−0.000 (0.014)	0.020 (0.012)
Candidate's gender: Female x RLS: high	−0.022 (0.018)	−0.023 (0.019)	0.022 (0.017)	−0.023 (0.018)	0.027 (0.020)	−0.007 (0.019)	−0.002 (0.016)
Candidate's gender: Female x GCO: male	−0.024 (0.021)	−0.036 (0.021)	0.007 (0.021)	−0.025 (0.021)	−0.029 (0.022)	−0.011 (0.021)	−0.051** (0.019)
Observations							1,449

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression models outlined in Subsection 4.3. More concretely, in each column, we present the coefficient estimates of the regression where we regress one of the stereotypes or attitude scales on (i) all candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics, while including all interaction terms (i.e., interactions between the gender of the job applicant and (i) all other candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table 6. Moderation analysis with stereotype and attitude scales as outcome variable (Continued).

	Openness	Respect towards authority	Sense of responsibility	Absence due to family responsibilities	Willingness to collaborate: employers	Willingness to collaborate: co-workers	Willingness to collaborate: customers
	(15)	(16)	(17)	(18)	(19)	(20)	(21)
A. CANDIDATE CHARACTERISTICS							
Gender: female	−0.009 (0.203)	0.333** (0.200)	0.095 (0.186)	0.478*** (0.178)	0.184 (0.204)	0.228 (0.188)	0.094 (0.184)
Race: black	0.121** (0.066)	0.019 (0.065)	0.081 (0.063)	−0.057 (0.062)	0.027 (0.058)	0.091* (0.060)	0.118** (0.065)
Commuting distance: more than 50 miles (CD)	−0.051 (0.070)	−0.019 (0.074)	−0.040 (0.073)	−0.041 (0.069)	−0.089* (0.069)	−0.162*** (0.069)	−0.115* (0.069)
Experience in the occupation: none	−0.552*** (0.080)	−0.567*** (0.074)	−0.827*** (0.074)	−0.108 (0.063)	−0.694*** (0.074)	−0.705*** (0.076)	−0.602*** (0.075)
Recent employment gap (EG)							
Due to unemployment	−0.332*** (0.086)	−0.272*** (0.084)	−0.428*** (0.083)	0.125 (0.076)	−0.216** (0.081)	−0.174 (0.084)	−0.366*** (0.085)
Due to health issues	−0.238*** (0.081)	−0.042 (0.083)	−0.221** (0.078)	0.123 (0.083)	−0.183* (0.082)	−0.106 (0.083)	−0.237** (0.086)
Due to family responsibilities	0.060 (0.081)	0.023 (0.084)	−0.137 (0.081)	1.051*** (0.091)	0.071* (0.077)	0.031 (0.081)	0.009 (0.080)
None (reference)	-	-	-	-	-	-	-
Extracurricular activities (ECA)							
Volunteering	0.278*** (0.091)	0.265*** (0.092)	0.270*** (0.086)	−0.010 (0.087)	0.262*** (0.089)	0.388*** (0.088)	0.378*** (0.091)
Sport activities	0.195** (0.080)	0.228** (0.081)	0.177* (0.082)	−0.095 (0.075)	0.217** (0.078)	0.336*** (0.085)	0.258** (0.082)
Cultural activities	0.240** (0.097)	0.142 (0.096)	0.074 (0.096)	−0.123 (0.087)	0.157 (0.092)	0.283*** (0.093)	0.237* (0.092)
None (reference)	-	-	-	-	-	-	-
Candidate’s gender: Female x Race: black	−0.167* (0.092)	−0.142* (0.087)	−0.137 (0.083)	0.026 (0.089)	−0.104 (0.083)	−0.157 (0.085)	−0.128 (0.091)
Candidate’s gender: Female x CD: more than 50 miles	−0.082 (0.104)	0.024 (0.103)	−0.074 (0.103)	0.078 (0.100)	−0.041 (0.108)	0.058 (0.102)	0.018 (0.104)
Candidate’s gender: Female x Experience: none	0.102* (0.124)	0.030 (0.121)	0.109* (0.113)	0.125 (0.092)	0.013 (0.122)	0.006 (0.119)	0.002 (0.121)
Candidate’s gender: Female x EG: due to unemployment	0.324* (0.140)	0.209 (0.134)	0.290 (0.123)	−0.129 (0.125)	0.063 (0.134)	0.097 (0.130)	0.359** (0.132)
Candidate’s gender: Female x EG: due to health issues	0.155 (0.126)	−0.017 (0.130)	−0.005 (0.119)	0.018 (0.128)	−0.033 (0.129)	−0.001 (0.125)	0.138 (0.127)
Candidate’s gender: Female x EG: due to family responsibilities	−0.187** (0.129)	−0.178** (0.130)	−0.081* (0.117)	−0.277** (0.126)	−0.354*** (0.126)	−0.284*** (0.126)	−0.203** (0.126)
Candidate’s gender: Female x ECA: Volunteering	0.029 (0.147)	−0.202 (0.144)	−0.095 (0.133)	−0.110 (0.132)	−0.000 (0.144)	−0.045 (0.134)	−0.110 (0.140)
Candidate’s gender: Female x ECA: Sport activities	0.052 (0.126)	−0.162 (0.131)	0.070 (0.125)	0.114 (0.127)	0.054 (0.131)	0.026 (0.135)	0.048 (0.133)
Candidate’s gender: Female x ECA: Cultural activities	0.048 (0.154)	−0.149 (0.150)	0.108 (0.149)	0.203 (0.144)	−0.012 (0.149)	−0.020 (0.151)	−0.050 (0.150)
B. APPLICANT POOL CHARACTERISTICS							
Gender composition applicant pool: majority is male (GCAP)	−0.009 (0.084)	−0.032 (0.088)	0.019 (0.081)	0.021 (0.081)	0.021 (0.084)	0.049 (0.083)	−0.049 (0.085)
Candidate’s gender: Female x GCAP: majority is male	0.007 (0.086)	−0.033 (0.081)	−0.068 (0.082)	−0.120 (0.098)	0.007 (0.080)	−0.074 (0.079)	0.081 (0.076)
C. JOB CHARACTERISTICS							

Level of required skills in occupation: high (RS)	−0.009 (0.017)	0.005 (0.018)	0.002 (0.016)	0.036** (0.016)	−0.005 (0.017)	−0.008 (0.017)	0.003 (0.018)
Level of required customer contact in occupation: high (RCC)	−0.023* (0.014)	−0.020 (0.014)	−0.014 (0.012)	0.024** (0.012)	−0.002 (0.013)	−0.014 (0.013)	−0.007 (0.013)
Level of required physical effort in occupation: high (RPE)	−0.004 (0.014)	0.007 (0.015)	−0.002 (0.013)	0.035** (0.014)	−0.006 (0.014)	−0.001 (0.014)	0.002 (0.014)
Level of required leadership skills in occupation: high (RLS)	0.015 (0.018)	0.002 (0.021)	−0.004 (0.016)	−0.003 (0.018)	0.004 (0.019)	0.016 (0.019)	0.012 (0.020)
Gender composition in occupation: male (GCO)	0.036 (0.023)	0.038 (0.025)	0.029 (0.021)	0.007 (0.018)	0.028 (0.022)	0.029 (0.022)	0.034 (0.022)
Candidate's gender: Female x RS: high	0.029 (0.018)	−0.005 (0.017)	−0.002 (0.016)	−0.028 (0.018)	0.037 (0.017)	0.029 (0.016)	0.017 (0.016)
Candidate's gender: Female x RCC: high	0.014 (0.016)	−0.005 (0.015)	−0.004 (0.013)	−0.012 (0.014)	−0.014 (0.014)	−0.007 (0.014)	−0.006 (0.013)
Candidate's gender: Female x RPE: high	0.016 (0.015)	−0.007 (0.015)	0.014 (0.013)	0.013 (0.015)	0.017 (0.014)	0.019 (0.013)	0.011 (0.012)
Candidate's gender: Female x RLS: high	0.002 (0.020)	0.014 (0.018)	0.016 (0.016)	0.002 (0.019)	−0.018 (0.018)	−0.007 (0.018)	0.000 (0.018)
Candidate's gender: Female x GCO: male	−0.040 (0.023)	−0.039* (0.023)	−0.028 (0.021)	−0.007 (0.021)	−0.022 (0.021)	−0.035 (0.020)	−0.028 (0.021)
Observations							1,449

Notes: The presented statistics are coefficient estimates and standard errors in parentheses for the regression models outlined in Subsection 4.3. More concretely, in each column, we present the coefficient estimates of the regression where we regress one of the stereotypes or attitude scales on (i) all candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics, while including all interaction terms (i.e., interactions between the gender of the job applicant and (i) all other candidate characteristics, (ii) the applicant pool characteristics, and (iii) the job characteristics. Standard errors are corrected for the clustering of the observations at the participant level. *** (**) ((*)) indicates significance at the 1% (5%) ((10%)) significance level.

Table A–1. Overview of names used in the experimental materials.

Associated race of names	Associated gender of names	
	Male	Female
White names	Neil Morrison	Molly Rose
	Steven Russell	Meredith Rogers
	Scott Sullivan	Katie Burns
	Matthew Owen	Allison Baker
	Luke Kelly	Kristen Pierce
	Jake Ryan	Sarah Miller
	Brad Richards	Megan Stone
	Paul Bennett	Emily Hart
Black names	Darius Mosby	Tyra Cooks
	Jermaine Jackson	Janae Washington
	Denzel Gaines	Keisha Towns
	Darnell Dawkins	Tamika Battle

Note: We give an overview of all 24 combinations of first and last names used in the experiment through which we signalled both an applicant's gender and race, as discussed in Subsection 2.1 and Subsection 2.3. As mentioned, all names were tested in a preliminary name categorisation experiment in which participants had to categorise all names in different (i) gender, (ii) race, (iii) social class, (iv) religious, and (v) immigrant generation categories. All presented names signalled the intended gender, race, social class, religious affiliation, and immigrant generation.

Table A–2. Jobs and corresponding job characteristics used in the experimental materials.

Job	Required skills	Required level of customer contact	Required level of physical effort	Required leadership skills	Gender dominance
Ophthalmic laboratory technician	Low	Low	Low	Low	No
Welder	Low	Low	Low	Low	Male
Payroll and timekeeping clerk	Low	Low	Low	Low	Female
Production, planning, and expediting clerk	Low	Low	Low	High	No
Door-to-door sales worker	Low	High	Low	Low	No
Food-batch-maker	Low	Low	High	Low	No
Operations research analyst	High	Low	Low	Low	No
Chemical engineer	High	Low	Low	Low	Male
Proof-reader	High	Low	Low	Low	Female
Human resources manager	High	Low	Low	High	No
Insurance sales agent	High	High	Low	Low	No
Wind farm support specialist	High	Low	High	Low	No

Note: Jobs were selected and categorised based on data provided by O*NET and IPUMS, as described in Subsection 2.2.

Table A–3. Job descriptions used in the experimental materials.

Job function	Job description
Ophthalmic laboratory technician	'The ophthalmic laboratory technician is responsible for cutting, grinding, and polishing eyeglasses and contact lenses. Additionally, he or she is responsible for assembling and mounting of lenses into frames'
Welder	'The welder is responsible for brazing or soldering together different components to assemble fabricated metal parts, using soldering iron, a torch, or a welding machine and flux.'
Payroll and timekeeping clerk	'The payroll and timekeeping clerk is responsible for compiling and recording employee time and payroll data. His or her duties include computing employees' time worked, the number of goods produced, and the number of commissions realised. Additionally, he or she is responsible for computing and posting wages and deductions or preparing paychecks.'
Production, planning, and expediting clerk	'The production, planning, and expediting clerk is responsible for coordinating and expediting the flow of work and materials within or between departments of the establishment according to the production schedule. His or her duties include reviewing and distributing production, work, and shipment schedules, conferring with department supervisors to determine the progress of work and completion dates, and compiling reports on inventory levels, costs, production problems, and the progress of work.'
Door-to-door sales worker	'The door-to-door sales worker is responsible for selling our goods and services door-to-door.'
Food-batch-maker	'The food batch-maker is responsible for setting up and operating equipment that mixes or blends ingredients used in the manufacturing of food products, including candy makers and cheese makers.'
Operations research analyst	'The operations research analyst is responsible for formulating and applying mathematical modelling and other optimising methods to develop and interpret information that assists management with decision making, policy formulation, and other managerial functions. His or her duties include collecting and analysing data, developing decision support software, service, or products, and developing and supplying optimal time, cost, or logistics networks for program evaluation, review, or implementation.'
Chemical engineer	'The chemical engineer is responsible for designing chemical plant equipment and devise processes for manufacturing chemicals and products, such as gasoline, synthetic rubber, plastics, detergents, cement, paper, and pulp, by applying principles and technology of chemistry, physics, and engineering.'
Proof-reader	'The proof-reader is responsible for reading transcripts and detecting and marking any grammatical, typographical, or compositional errors in this transcript'
Human resources manager	'The human resources manager is responsible for planning, directing, and coordinating human resources activities and the staff of the organization.'
Insurance sales agent	'The insurance sales agent is responsible for selling life, property, casualty, health, automotive, or other types of insurance.'
Wind farm support specialist	'The wind farm support specialist is responsible for inspecting, diagnosing, adjusting, and repairing wind turbines. His or her duties include performing maintenance on wind turbine equipment including, resolving highly technical and complex electrical, mechanical, and hydraulic malfunctions.'

Note: Job functions and descriptions were provided by O*NET, as described in Subsection 2.2.

Table A–4. Comparison between participant characteristics and characteristics of HR professionals in ACS.

	(1)	(2)
Participant characteristics	Mean among participants in experiment	Mean among HR professionals in ACS
Gender: female	0.435	0.670
Age	40.711	45.363
Race: only white	0.820	0.811
Highest educational degree: secondary education or lower	0.150	0.140
Highest educational degree: tertiary education	0.850	0.860

Notes: We combined ACS data conducted in the years 2010-2019 and selected all respondents with occ2010 occupation codes 0130 (Human Resources Managers), 0620 (Human Resources, Training, and Labour Relations Specialists), and 5360 (Human Resources Assistants, Except Payroll and Timekeeping).