Experimentel Evidence of Discrimination in the Labour Market
Intersections between ethnicity, gender, and socioeconomic status
Dahl, Malte; Krog, Niels

Publication date:
2017

Document version
Publisher’s PDF, also known as Version of record

Citation for published version (APA):
Experimentel Evidence of Discrimination in the Labour Market
Jacobsen, Malte Dahl; Krog, Niels

Publication date: 2017

Citation for published version (APA):
Experimental Evidence of Discrimination in the Labour Market: Intersections between ethnicity, gender, and socioeconomic status

WORKING PAPER
Presented at the ASA Annual meeting 2017

Malte Rokkjaer Dahl and Niels Soegaard Krog

Abstract

This paper presents evidence of ethnic discrimination in the recruitment process from a field experiment conducted in the Danish labour market. In a correspondence experiment, fictitious job applications were randomly assigned either a Danish or Middle Eastern-sounding name and sent to real job openings. In addition to providing evidence on the extent of ethnic discrimination in the Danish labour market, the study offers two novel contributions to the literature more generally. First, because a majority of European correspondence experiments have relied solely on applications with male aliases, there is limited evidence on the way gender and ethnicity interact across different occupations. By randomly assigning gender and ethnicity, this study suggests that ethnic discrimination is strongly moderated by gender: minority males are consistently subject to a much larger degree of discrimination than minority females across different types of occupations. Second, this study addresses a key critique of previous correspondence experiments by examining the potential confounding effect of socioeconomic status related to the names used to represent distinct ethnic groups. The results support the notion that differences in call-backs are caused exclusively by the ethnic traits.

1 The authors would like to thank Peter Thisted Dinesen, Martin Vinas Larsen, and Jens Olav Dahlgaard at University of Copenhagen for valuable advices. The authors also thank participants at the American Sociological Associations yearly meeting, 2017, participants at the annual meeting at the Danish Political Science Association, 2016, and Florian Foss for comments on this paper. The replication data and code for conducting the data analysis is available upon request.
1. Introduction

In many European countries, non-western immigrants face substantial employment deficits and wage differentials. This has grave consequences for the unemployed individuals and for the societies they inhabit, and has given rise to intense political debates about the rights and obligations of ethnic minorities (Dancygier & Laitin, 2014). There are a multitude of possible explanations for such labour market differentials, but several studies suggest that discrimination in the hiring process is an entry barrier to ethnic minorities (Baert, Cockx, Gheyle, & Vandamme, 2015; Birkelund, Heggebø, & Rogstad, 2017; Carlsson & Rooth, 2012; Kaas & Manger, 2012; Pager, Western, & Bonikowski, 2009). A range of designs have been leveraged to study labour market discrimination, but field experiments are seen as a significant methodological advance (Bertrand & Duflo, 2016; Neumark, 2016). So-called correspondence experiments, where applications are submitted to job advertisements in the name of fictitious applicants, is an increasingly common tool for social science researchers. These experiments enable researchers to identify if, and to what extent, group affiliations—signified by names—affect the chances of getting a call-back.

Despite the fact that an increasing number of field experiments have examined labour market discrimination, important questions remain unanswered. Firstly, in a European context, many correspondence studies have refrained from randomly assigning both ethnicity and gender, and therefore have overlooked the possible intersections between the two. By only using male applicants and assuming effect homogeneity across gender, many previous studies might not have told an accurate story about the overall ethnic disparities. If anything, the evidence from studies that do manipulate both traits points towards larger ethnic penalties among male applicants. Recent studies suggest that the interaction between gender and ethnicity is highly dependent on the composition of jobs
included in the experiment, e.g. the share of private sector jobs or the gender-balance in the occupations included (Bursell, 2014; Midtbøen, 2016). Hence, considering varying effect sizes across gender and across different occupations is essential to ensure generalisability when studying ethnic or racial discrimination. Furthermore, if such intersection between ethnicity and gender exists, it poses an important question of why members of the same ethnic group face different outcomes.

Secondly, correspondence experiments face a challenge related to the internal validity when using names to manipulate characteristics of interest. Names do not exclusively signify racial or ethnic affiliation but contain a bundle of information. For example, the applicants’ socioeconomic status (SES) might be inferred from their names, and if the popular majority names used in correspondence experiments are perceived as having a higher socioeconomic status than the distinct ethnic or racial minority names, it confounds the ethnic trait (Bertrand & Mullainathan, 2004; Fryer & Levitt, 2004). In other words, if popular majority and ethnic minority names is also a comparison across SES, it violates the excludability assumption that the effect is caused solely by the ethnic trait (Butler & Homola, 2017). Ultimately, this implies that it is not known whether ethnic or SES discrimination causes the ethnic disparities found in previous experiments.

To address these questions, we conducted a field experiment in which 800 generic applications were sent to job openings in the Danish labour market. Each job opening received two equally qualified applications which were randomly assigned either a traditional Danish-sounding name or a Middle Eastern-sounding name. Applicants’ ethnicity and gender were randomly assigned in order to study the interaction between applicant gender and ethnicity. Furthermore, to examine whether the socioeconomic status of applicants’ names confounds the effect of ethnicity, half of the majority names
were randomly assigned from a pool of names associated with low SES. The pool of jobs applied for was diverse, representing various skill sets and spanning both the public and the private sector. The applications contained relevant experience and education as well as markers signalling that the applicants were competent, tolerant and likeable people. In other words, the applicants were highly qualified for the jobs they applied for, which, according to previous empirical studies, is likely to minimise the employers’ incentive to discriminate (Agerström et al., 2012; Birkelund et al., 2017; Kaas & Manger, 2012).

The paper reports three main findings. First, similar to other studies, we find evidence of considerable discrimination in the hiring process with a call-back ratio of 1.52 for job interviews between applicants with traditional Danish-sounding names and Middle Eastern-sounding names. This means that applicants with Middle Eastern-sounding names on average have to apply for 52% more jobs to receive the same number of call-backs as applicants with Danish-sounding names. Second, the results show that discrimination varies substantially by gender, with a larger ethnic difference among male applicants. The interaction between ethnicity and gender exists across sector and in gender-balanced occupations as well as in occupations that are highly gender-dominated. Third, we find no evidence that the socioeconomic status of the majority applicants’ names moderates the effect of the ethnic treatment, which supports the notion that the differences between treatment groups are caused by the ethnic trait and are not associated with socioeconomic status.

The paper proceeds as follows: Section 2 briefly introduces the Danish context, while Section 3 presents the theoretical backdrop and existing evidence. Section 4 derives the hypotheses and section 5 presents the experimental design in detail. Section 6 presents the results, while section 7 concludes and discusses the findings.
2. The Danish context

The motivation for examining discrimination against workers with Middle Eastern-sounding names is twofold. First, there has been a rapid increase in the number of non-Western immigrants and descendants in Denmark in the last 30 years, most of whom are of Middle Eastern origin, with the largest groups being from Turkey, Lebanon, Pakistan and Iraq. The immigration from non-Western countries was originally intended to supply the booming labour market in the late 1960s, but since the mid-1970s most of the migration has consisted of asylum seekers and family reunifications. Today, non-Western immigrants and descendants is a significantly larger group than immigrants and descendants from Western countries and comprises approximately 8.5 percent of the total Danish population (Statistics Denmark, 2017). Second, immigrants and descendants of Middle Eastern origin has been and still is a very salient group in the persistent political debate over immigration and integration in Denmark (Simonsen, 2017). Discussions about the economic and cultural integration of immigrants with Middle Eastern origin revolve around the comparatively high unemployment rates and the fact that minorities of Middle Eastern origin fare worse on a number of socioeconomic indicators such as education, income-levels and crime rates (Statistics Denmark, 2017). This is reflected in attitudes among native Danish citizens who generally recognise immigration and integration as vital political topics. Anti-immigrant attitudes in Denmark are equivalent to most other European countries (Dinesen & Sonderskov, 2015), which also manifests in support for parties running on a sceptical immigration platform (Mudde, 2013; Rydgren, 2008). Finally, research on experienced discrimination in the Danish labour market shows that large shares of non-western immigrants have experienced labour market
discrimination (Jensen et al., 2013) and that employers perceive language and cultural issues as barriers when hiring ethnic minorities (Slot, 2008).

3. Existing evidence and theoretical backdrop

A wealth of research has examined attitudes towards ethnic minorities and self-reported experiences of discrimination, but since discrimination is a sensitive topic and events can be misjudged or overlooked, it remains unclear to what extent experiences of discrimination correspond to a reliable representation of reality (Pager & Shepherd, 2008). However, correspondence experiments makes seemingly ‘immutable characteristics’ manipulable by exposing units to signals of the given characteristic (Sen & Wasow, 2016). The fundamental idea is to hold constant anything but the group characteristic being examined. By exposing employers to randomly assigned traits associated with a given group—e.g. gender, race or ethnic categories—any difference in outcomes can be ascribed to the treatment. Conducting the experiments in the field is key to measuring actual behaviour when studying a highly sensitive topic such as discrimination.

Researchers across a large number of countries have devoted considerable effort to accumulating evidence that corroborates the notion that discrimination of outgroups exists in labour markets. In a review of correspondence experiments conducted in the period 1990-2015, Zschirnt & Ruedin (2016) conclude that experimental research consistently finds proof of ethnic or racial discrimination in the hiring process across OECD countries. It should be kept in mind that results from different correspondence experiments cannot

---

2 The method has also been applied in a number of domains outside the labour market, measuring differential treatment in housing markets (Fang, Guess, & Humphreys, 2015), in the market place (Ayres, & Siegelman, 1995), in the sharing economy (Edelman, Luca, & Svirsky, 2017) or state legislators’ responsiveness to requests from voters (Butler & Broockman, 2011).
be compared directly, since variations in experimental designs and the demand for labour in local contexts influence findings. Nevertheless, when averaging across studies, minorities have to apply for 50% more jobs to receive the same number of job interviews as the majority group (Zschirnt & Ruedin, 2016). In summary, these differences are consistent and substantial across a large number of countries.

**Theories of discrimination**

Following the empirical evidence of differential treatment, the question of why ethnic or racial discrimination occurs is obviously of immense interest. Two theories of discrimination dominate the literature. In the *taste-based discrimination* model introduced by Becker (1958), discrimination is seen as the result of an irrational distaste towards certain groups. In other words, because of prejudiced employers, co-workers or customers, there is a disamenity value to employing minority workers, resulting in preferential hiring and wage differentials (Guryan & Charles, 2013). An alternative explanation for discrimination is found in the so-called *statistical discrimination* models (Phelps, 1972). The foundation of these models is that employers have limited information about applicants’ productivity, giving them an incentive to utilise their knowledge on the average productivity of the applicants’ group in the evaluation of individual applicants. Thus, if ethnicity correlates with undesired traits, discrimination becomes an optimisation strategy (Arrow, 1998).

While the theoretical premises of the two models of discrimination are fundamentally different, it has proven difficult to empirically differentiate between them (Bertrand & Mullainathan, 2004; Dancygier & Laitin, 2014). First, and most fundamental, it is difficult to elicit distinct observable implications between the two theories, thus making it
difficult to distinguish between them empirically. Secondly, different types of
discrimination might interact over time. Disadvantages initially caused by taste-based
discrimination can eventually initiate real group differences in education or labour market
outcomes, creating a basis for statistical discrimination. Even if one type of
discrimination is successfully identified at one point in time, it can be a complex result of
preceding processes (Dancygier & Laitin, 2014). With these limitations in mind, well-
designed correspondence experiments enable researchers to identify which components
trigger discrimination (Sen & Wasow, 2016). When the particular ways in which group
characteristics, alone and in combination, result in disparities is understood,
discrimination is much more likely to be recognised and addressed.

*Intersections between ethnicity and gender in the hiring process*

While the combination of gender and ethnicity in labour market discrimination has
been under-researched in the field experimental literature, the general literature on
intersections between gender and ethnicity is vast. Two of the most prominent notions
about gendered ethnic discrimination outline very different empirical implications; first,
from the perspective of intersectionality it is argued that minority women will suffer the
largest disadvantage since they occupy the lowest position in both social categories, being
female and belonging to an ethnic minority group (Harnois, 2015; Ransford, 1980). This
dual oppressive system—whether it is the notion of additive jeopardies (Beal, 1970) or
multiplicative jeopardies (King, 1988)—can be translated into a ‘supplementary
discrimination hypothesis’ that expects a larger ethnic gap among female applicants.

---

3 One way of assessing the implications of theories of discrimination has been to examine heterogeneous treatment effects—e.g.,
varying effect sizes across firm size or customer contact. While such treatment-by-covariates effects can be interesting, they are not
solid answers to causal questions. Jobs with certain characteristics might differ systematically on a number of unobserved variables
that alter the explanation
An alternative prediction is found in the social psychology literature. Social dominance theory, a general model of hierarchically structured relationships among social groups, argues that ethnic conflict is primarily executed by and targeted against males (Sidanius & Pratto, 2001). It is claimed that minority males are perceived as a greater threat and therefore are the primary target of discrimination, while outgroup females are less susceptible to discrimination. This is also known as the ‘Outgroup-Male-Target hypothesis’ (Navarrete, McDonald, Molina, & Sidanius, 2010). Arguably, both notions can be understood along the lines of taste-based discrimination, with a focus on factors unrelated to the productivity-related characteristics. However, interactions between ethnicity and gender can be a result of statistical discrimination too. For example, minority males in general have lower educational credentials than minority females and they are overrepresented in the criminal justice system, which can be expected to translate into different levels of discrimination.

Although the majority of European correspondence experiments rely on male applicants, there is good reason to consider the importance of gender in studies on discrimination. Labour market field experiments that manipulate applicants’ gender generally support the Outgroup-Male-Target hypothesis (Andriessen et al., 2012; Arai et al., 2016; Liebkind et al., 2016; Midtbøen, 2016) although this finding is not consistent. Some correspondence experiments find little or no variation in ethnic discrimination across gender (Birkelund et al., 2017; Blommaert et al., 2014; Derous et al., 2012). Moreover, one line of research emphasises how intersections between ethnicity and gender differs substantially across the occupations included in the sample. In a study conducted in the Swedish labour market, Bursell (2014) finds a larger ethnic difference among males in male-dominated occupations, while the ethnic difference is largest among
females in female-dominated occupations. In the Norwegian labour market, Midtbøen (2016) similarly finds that the effect of the ethnic trait is larger among male applicants, but not in gender-integrated occupations in the private sector. This is especially important since most of the aforementioned studies adjust the research design according to patterns in gender stratification such that job openings within occupations that were very male-dominated only received applications by male applicants and vice versa. In this study, we randomly assigned pairs of applicants with the same gender to each job, which allows us to compare call-back rates across ethnicity and gender without any adjustments to the types of job applied for. In addition, we can assess effect heterogeneity by breaking down the results across occupations that are highly stratified by gender.

Using names as proxies for ethnicity and potential confounding

Correspondence studies examining ethnic discrimination rely on the assumption that differences in call-backs are exclusively due to the signal that the name provides about ethnicity. However, perceptions about names might be influenced not only by the population racial/ethnic composition of a name, but also its population socioeconomic status (Gaddis, 2017). This becomes an issue in correspondence studies where distinct minority names might be perceived as low-status compared to the distinct majority names. Hence, it is possible that studies relying on racial or ethnic distinct names are picking up a confounding relationship between ethnicity/race and SES (Fryer & Levitt, 2004). Thus far, this notion is largely theoretical and the evidence supporting it is limited.

---

4 One argument for doing so is to avoid evoking suspicion among employers in gender-stratified occupations where two applications from similar qualified applicants of the same gender will seem odd (Arai et al., 2016). Another argument relates to the real life consequences of discrimination: if very few female candidates work in construction, the need to examine and address discrimination against female candidates is arguably smaller.
and ambiguous. In an American context, two recent correspondence experiments in the labour market did not replicate the previous results of racial disparities (Darolia et al., 2016; Deming et al., 2016). One possible reason for this is that the names used to signal race had different connotations of SES than previous correspondence experiments. On the other hand, the divergence from previous research could also be due to employers not being able to precisely identify applicants’ race, which attenuates the effect.

In a European context, Jackson (2009) conducted a field experiment in the UK to examine if different traits signifying social class had an effect on call-backs from employers. Overall, the combination of different high-status characteristics only resulted in small advantages, but the largest observed (positive) effect of an individual treatment, although only borderline significant, was attributed to holding an elite name. Hence, there is good reason to examine potential effects of socioeconomic status related to names to obtain a valid measure of ethnic discrimination and to provide new information for the discussion of the importance of SES.

Since Middle Eastern-minorities in Denmark fare worse on a number of socioeconomic indicators compared to native Danes, it is plausible that employers perceive them as a low-status group. One way to address this in the research design is to match majority and minority applicants on social class (Gaddis, 2015). However, since there is no reason to believe that employers are able to distinguish high and low-SES minority names, it is only possible to manipulate the SES-component among majority

---

5 Butler & Homola (2017) perform an ex post analysis of an audit study on political responsiveness using public records to assess the importance of socioeconomic status and political resources reflected in names. They find no evidence that these signals predict the probability of legislators’ likelihood of responding.

6 To circumvent the risk of ascertaining socioeconomic status of racially distinct first names, Darolia et al., (2016) only signify race through distinct surnames, which might not be a clear signal of race.
applicants. The section on the study’s research design outlines the strategy for choosing the specific names.

4. Hypotheses

We examine three hypotheses that were all pre-registered at the EGAP.org database. In answering these hypotheses, we rely on the framework outlined by Sen & Wasow (2016), conceptualising ethnicity as a composite variable rather than a single uniform entity. In this framework, ethnicity is a fusion of several factors, such as region of ancestry, religion, or socioeconomic status, which might trigger discrimination in different ways. By exposing employers to different manipulations of randomly assigned characteristics, we can disentangle how different aspects alone and in combination affect behaviour. This is a useful methodological starting point that enables a deeper understanding of which and when social groups are subject to differential treatment.

First, we are interested in the overall difference, ceteris paribus, between the two ethnic groups. With the consistent findings of ethnic discrimination in previous European correspondence experiments in mind, we hypothesise that there will be an overall significant difference between the majority group and the ethnic minority group:

H1: Applicants with Middle Eastern-sounding names are less likely to receive a call-back compared to applicants with traditional Danish-sounding names.

Since perceptions of males and females within the same ethnic group might differ, it is essential to include both groups in the treatment in order to get a general measure of

---

7 We initially assumed that the large majority of employers would belong to the majority group. From the names of employers or HR-managers that we contacted, we only identified one with a minority name.

8 See details at appendix H.
ethnic discrimination. Following the theoretical and empirical work outlined in the previous section, we test the gendered nature of ethnic discrimination in a second hypothesis:

H2: The difference in call-backs between majority and minority applicants is larger among male applicants than among female applicants.

Finally, we examine if the SES of the names used to signal ethnicity might be a confounding variable by manipulating the SES of the majority applicants’ names.

H3: The difference in call-backs between majority and minority applicants is larger when the majority applicant holds a popular name than when the majority applicant holds a low socioeconomic status name.

5. Experimental design and implementation

The experimental design in correspondence studies, especially the quality of the applications and the jobs applied for, can influence results substantially (Neumark, 2012). In this study, we aimed for a design that minimises employers’ incentive to discriminate against the minority applicants. These design features and the implementation of the experiment are described in detail in the following sections.

Treatments and randomisation details

We randomised the assignment of three different treatments – ethnicity, gender and socioeconomic status – using names as proxies. The applicants’ gender was also explicitly stated in the CVs to avoid potential misconceptions of the treatment. Each job received two applications, one with a Danish-sounding name and one with a Middle Eastern-sounding name. Gender was randomly assigned pairwise, so the applicants for any given
job were either two males or two females. In other words, we randomised ethnicity *within* and gender *across* the job ads (Fig. 1). Finally, in order to study if SES affects chances of a call-back, the traditional Danish-sounding names were randomly assigned from two pools: either the most popular Danish names or Danish low-status names.

Presumably, it is difficult for most employers to differentiate the socioeconomic status of various Middle Eastern-sounding names, which is why we only manipulated the SES of the pool of Danish-sounding names. We can test the importance of SES by comparing, on the one hand, the difference in call-backs between minority applicants and applicants with the most popular Danish-sounding names, and, on the other hand, the difference in call-backs between minority applicants and applicants with low-status Danish-sounding names.

**Figure 1. Assignment to treatment and control groups**

![Figure 1. Assignment to treatment and control groups](image)

It is essential that the names serve as distinct signals of ethnicity and gender (Riach & Rich, 2002). Since the most common Middle Eastern-sounding names are quite distinct, it can be assumed that employers are able to differ between the ethnic traits. We used a large number of different names (stimuli sampling) to ensure that effects are caused by the categories of interest and not the unique characteristics of a specific name (Wells &
The names can be divided into three groups: a pool of the most common Danish-sounding male and female first names, a pool of the most common male and female first names used by Middle Eastern minorities in Denmark (Statistics Denmark, 2015) and a pool of Danish male and female low-status names. Furthermore, two pools of surnames with the most common traditional Danish-sounding (Middle Eastern-sounding) surnames were randomly paired with the pool of Danish-sounding (Middle Eastern-sounding) first names.

The Middle Eastern-sounding names are the most common minority names and hence represent a large share of the minority population. All of the minority names figure in a dictionary of Muslim first names that are often used in Denmark (Meldgaard, 2005). Furthermore, since all candidates have a Middle Eastern-sounding first name and a Middle Eastern-sounding last name, we effectively rule out any confusion about the precision of the ethnic signal.

It might be expected that some minority names evoke stronger Muslim connotations (e.g. Mohammed) that result in larger ethnic penalties, but call-backs are evenly distributed across names (Appendix J). To identify the pool of low SES-names, we used a dataset on the social characteristics of Danish citizens. We constructed an index of SES by using data on average income, crime rates and unemployment for Danish citizens with a given name. Furthermore, we excluded first names with an average age below 20 years or above 40 years. Hence, we identified the low-SES names from the bottom decile of the index and chose American-inspired names that fit a common Danish stereotype of being low-status (See details in appendix A).

---

9 We used data on the most common Danish names from Statistics Denmark, (2015). See a list of names in Appendix A.
10 The dataset included register data on more than 3.8 million Danes.
Constructing applications and CVs

The applications were designed to meet three criteria: i) they should be perceived as real applications, ii) each application in a pair should be different to avoid arousing suspicion and iii) the applicants should be equally qualified (Midtbøen & Rogstad, 2012). In the Danish labour market it is standard procedure to write a one-page application letter and attach a CV with the inclusion of a phone number and email. We constructed the generic applications building on four paragraphs: introduction, motivation, experience and personal interests. For each paragraph, we constructed two slightly different texts (A and B). By utilising all possible combinations of the A and B paragraphs, we got 16 unique pairs of applications where each pair consisted of each other’s opposite (e.g. one possible pair is A-A-B-B and B-B-A-A). We randomised the assignment of CVs to each application and randomised which application to send first (with 2-3 days between sending the applications for a given job). There were no differences in call-backs for job interviews across the different application pairs and CVs, which supports the notion that employers perceived them as equally qualified (Appendix B).

Table 1. Application templates

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction</strong></td>
<td>I believe the job as [] represents an excellent opportunity for me. I am highly motivated and have experience working as []. I am currently employed at [], but I recently moved to [] which is why I am looking for a job closer to my new home.</td>
<td>I would like to put forth my candidature for the position as []. I am very motivated and believe I have the experience to meet the requirements. I am currently employed as [] but after [] years I now feel it is time to look for new challenges.</td>
</tr>
<tr>
<td><strong>Motivation</strong></td>
<td>I enjoy working [in a team/on my own] and I will thrive in a position as []. I am familiar with [] and know you</td>
<td>Working [independently/ with good colleagues] motivates me, and I see the job at [] as a great match. [] is without doubt a</td>
</tr>
</tbody>
</table>
are among the most ambitious/most professional/best workplaces. I am very passionate about my work and would love to be part of a team with a professional and positive attitude.

**Experience**

I am a very ambitious [] and I work hard to achieve good results. In my previous jobs, I have always been given responsibility and I have been an esteemed employee. I hope that I can achieve the same in the job at []. From [year] to [year] I worked in [] with []. The last [] years I have worked in [] why I bring experience to the job.

I have always worked hard and with great dedication. As [] I have been a trusted employee with a lot of responsibility. In my jobs at [] and [] where I worked for [] and [] years, respectively, we managed to [increase sale/provide a great service/do a professional job each day]. I hope to continue in the job at [].

**Personal information**

I am a very positive and social person who believes good relationships in the workplace is important. In my spare time, I do a lot of sports and run several times a week together with good friends. Besides that, I love to travel and scuba dive with my boy/girlfriend.

I am an optimistic person with a good sense of humour. Having a good relationship with my colleagues is important to me.

I would appreciate the opportunity to meet with you and discuss how my qualifications can benefit your organisation [/company].

Sincerely,

[Name]

The applications were largely generic but held a few empty spaces that were completed in each specific case in order to match an application with a particular job. For example, the applicant states: ‘I enjoy working […]’ and we added ‘independently’ or ‘in a team’ depending on the given job. If the job ad explicitly requested specific requirements, these were incorporated into the CVs (e.g. software skills, a driver’s license etc.). The modification of the applications for each job secured a sufficient chance of receiving a call-back. It is impossible to know the relative quality of the applications,
since it depends on other applicants applying; however, the absolute quality can be adjusted (Neumark, 2012). While underqualified applications will receive few or no answers, over-qualified applications will rule out statistical discrimination by receiving a call-back too often. Broadly speaking, qualifications can be divided into cultural characteristics and professional characteristics. Our applicants have strong professional characteristics, possessing the necessary educational background and relevant experience from previous jobs.\textsuperscript{11} The cultural characteristics are signalled in the perfectly written Danish and the fact that applicants love to cook with their boyfriend/girlfriend, do triathlons or jog with good friends. These markers of a Danish middleclass lifestyle were included in order to reduce the perceived cultural differences between the applicants, hence minimising the incentive to discriminate against the minority applicant. Finally, in the CV it was stated that the applicant was born in 1984 and the gender of the applicant was indicated explicitly. In summary, the applicants were highly qualified for most of the jobs applied for, which is also indicated by the high call-back rates. Following previous studies, the high standard of the applications can be expected to reduce the ethnic discrimination (Kaas & Manger, 2012). As such, the study constitutes a least-likely case for observing discrimination.

\textit{The sample}

The empirical analysis builds upon experimental data collected between September 2015 and June 2016. In total, 800 applications were sent in response to 400 job openings. We\textsuperscript{11} The applicants’ CVs mentioned real educational institutions in the section on educational background and real workplaces in the section on experience to maximise realism. We did not receive any comments indicating that employers had been in contact with institutions or workplaces.
sampled the jobs from the online employment portal, Jobindex.dk. Jobindex is the largest employment portal in Denmark and covers various types of occupations.

To provide a comprehensive picture of potential discrimination, we applied for 23 different types of occupations within six different occupational categories: Office and administrative support, Education, Health care, Retail, Construction and extraction, and Marketing and sales (see all occupations in appendix D).

We maximised the geographical variation and applied for jobs from all five Danish regions. In total, 278 of the ads were private sector jobs. Hence, the sample comprises a broad geographical scope and covers a relatively large segment of the Danish labour market.

We excluded academic jobs as well as highly technical jobs from the sample since these would require detailed knowledge of essential skills, specific applications, recommendations and lengthy CVs that would exceed the generic applications used in this experiment. The sample includes occupations that are dominated by either women or men and occupations where the labour force is mixed. We coded all jobs according to the gender division in the specific occupation. Using a threshold of 20 percent, we identified 112 jobs as female-dominated and 90 jobs as male-dominated, while the remaining 198 jobs had at least 20 percent male and female workers ( appendix D). In order to minimise the risk of receiving an answer before sending the second application, we only applied for jobs where the expiration date from the job advertisement was two weeks or more. Some job ads demanded a picture of the candidate, in which case we did not apply.

Randomisation and handling of answers

For each job ad, we initially registered background information (sector, number of employees, language requirements and contact information) and adapted the application
and CV templates to the specific job before both applicants were finally assigned a gender and each application was assigned an ethnic affiliation. By finishing both applications first and randomly assigning names afterwards, we avoided the risk of unintentionally biasing the quality of the applications. If employers contacted one or both of the applicants, the job interview offer was politely declined.

We define a call-back as a personalised contact in the form of a message on the answering machine or an email from the potential employer with an invitation for a job interview. In a few cases, employers contacted applicants with additional questions or they asked for proof of education, which did not qualify as a call-back. All communication with employers were archived on either email or answering machines. In 178 out of 222 call-backs, we received an email.

**Ethical considerations**

There is a large body of literature on the ethical considerations involved when conducting correspondence experiments, including advice on how to minimise harm to subjects involved (see Riach & Rich (2004) and Zschirnt (2016) for an extensive discussion of ethics in correspondence experiments). Central arguments for the legitimacy of carrying out correspondence experiments include i) the question of discrimination is of high societal importance, ii) that there is no other way to credibly retrieve this vital information and iii) if the research is prepared and carried out carefully, there is only a very limited detrimental effect on the employers tested (Riach & Rich, 2004; Zschirnt, 2016).

On the last point, we conducted a pilot study to confirm that the experimental setup and the logistic of sending and handling the call-backs operated efficiently. Furthermore, in order to ensure sufficient statistical power without having to contact an excessively
large number of employers, we conducted a power analysis before the collection of data. Finally, we carefully considered how to minimise any inconveniences to the employers that were contacted as a part of the experiment. The main cost to employers is time, so we answered all requests as quickly as possible, explaining that the applicant had just found another job.\(^{12}\) We did not debrief employers in order to avoid the risk of making them doubt the credibility of future (minority) applicants. Furthermore, we analysed data on an aggregated form to ensure the anonymity of all individual employers in order to avoid associations between specific companies and this study (Pager, 2007).

**Balance check**

We performed a balance check to judge whether the random assignment procedure seems to be reliable (Gerber & Green, 2012). The main treatment, ethnicity, is necessarily balanced across covariates since all job ads received an application from both ethnic groups. However, this is not the case for gender and SES, and therefore we tested whether observed covariate imbalances are larger than expected from chance alone. To examine this, we regressed the treatment indicators (gender and SES respectively) on the available covariates\(^{13}\) and calculated the heteroscedasticity-robust Wald statistic for the hypothesis that all the coefficients on the covariates are zero (Lin, Green, & Coppock, 2016). In order to calculate the related p-values, we used randomisation inference to create a distribution of Wald statistics under the null hypothesis of no systematic imbalance. The results show no reason to reject the null hypothesis that the pre-treatment covariates are

\(^{12}\) We created four individual email addresses and set up four phone numbers with voice mails. Thus, we were able to send out applications and receive answers for all four combinations of gender and ethnicity. All invitations were either recorded from the answering machines or saved as screenshots from emails.

\(^{13}\) Covariates include Size (number of employees); Sector (public/private); Education (if education after high school was needed); Language required and customer contact.
not systematically related to the treatment (p-values: 0.43 and 0.92) and we therefore consider the assignment of treatments to be balanced (see appendix C for further details).

6. Results and interpretation

To recap, this experiment answers three main questions. First, are equally qualified applicants with either a Danish-sounding name or a middle Eastern-sounding name treated differently by employers? Second, is the effect of having a minority name moderated by applicants’ gender? Finally, is ethnic discrimination exclusively caused by the ethnic trait or does socioeconomic status confound the effect? All p-values and standard errors are obtained from randomization inference with 100.00 iterations.¹⁴

Main results

In total, 800 applications were sent to 400 jobs and at least one applicant received a call-back with an invitation for a job interview in 39.5% of these. Since each job opening received two equally qualified applications, we can observe two outcomes for each workplace. As is evident from Table 2, the call-backs were not equally distributed, with a substantial gap between minority and majority applicants. The majority applicants received a call-back rate of 33.5%, which is a substantively higher share than the minority applicants, who received a call-back on 22% of the applications. The difference corresponds to 11.5 percentage points (p < 0.01) and a ratio of 1.52, implying that minority applicants need to send 52% more applications to receive the same number of invitations as applicants with traditional Danish names. In the literature there are different ways to report outcomes from correspondence experiments. In this study, we first and

¹⁴ By reproducing the known randomisation procedure a large number of times the distribution of the test statistic under the sharp null hypothesis can be approximated with a high degree of precision (Gerber & Green, 2012)
foremost pay attention to the relative call-back ratio and the difference-in-means (DIM), but table 2 also reports the level of net discrimination, which is a common measure of discrimination in the literature. For all applicants, the net discrimination rate is 0.29, while it is 0.41 and 0.15 for male and female candidates respectively.\(^{15}\)

### Table 2. Distribution of Call-backs

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call-back majority</td>
<td>36.2</td>
<td>30.6</td>
<td>33.5</td>
</tr>
<tr>
<td>Call-back minority</td>
<td>19.3</td>
<td>24.9</td>
<td>22</td>
</tr>
<tr>
<td>Ratio</td>
<td>1.88</td>
<td>1.23</td>
<td>1.52</td>
</tr>
<tr>
<td>Difference-in means</td>
<td>16.9***</td>
<td>5.7*</td>
<td>11.5***</td>
</tr>
<tr>
<td></td>
<td>(3.66)</td>
<td>(3.15)</td>
<td>(2.41)</td>
</tr>
<tr>
<td>Neither invited</td>
<td>121</td>
<td>121</td>
<td>242</td>
</tr>
<tr>
<td>Both invited</td>
<td>29</td>
<td>35</td>
<td>64</td>
</tr>
<tr>
<td>Only majority invited</td>
<td>46</td>
<td>24</td>
<td>70</td>
</tr>
<tr>
<td>Only minority invited</td>
<td>11</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>Net discrimination</td>
<td>0.41</td>
<td>0.15</td>
<td>0.29</td>
</tr>
<tr>
<td>N</td>
<td>207</td>
<td>193</td>
<td>400</td>
</tr>
</tbody>
</table>

\(^*p < .1; \^{*\*}p < .05; \^{*\*\*}p < .01\)

Standard errors are in parentheses.

Table 2 also reports the DIM estimates and the related standard errors obtained from randomisation inference with 100,000 iterations under the sharp null hypothesis, assuming no effect of ethnicity for all applicants (appendix F). It is extremely unlikely (p < 0.001) that the overall ethnic difference in means of 11.5 percentage points would have occurred by chance. If we break down the results into occupational categories, we see that although the relative difference varies, the majority applicant is preferred over the

\(^{15}\) The measure of net discrimination treats cases with no call-backs as non-observations and is obtained by dividing the difference between observations where only the majority was invited and observations where only the minority was invited with the number of observations where at least one candidate was invited (see a discussion in Riach & Rich, 2002).
minority applicant in all six occupational categories (see appendix D for details). Overall, the results suggest that employers across occupations use ethnicity as an important decision-rule when evaluating applications, and hence that applicants with a Middle Eastern background are subject to discrimination.

*Call-backs across gender and ethnicity*

From the results reported in Table 2, it is noticeable that the ethnic difference in call-backs seems to be gender-reliant. The results are visualised in Figure 2 a), showing the difference in call-backs between majority and minority applicants for female applicants, male applicants, and all applicants respectively.

To test if the interaction between ethnicity and gender is significant, we regress a call-back dummy on ethnicity and gender of the applicant as well as the interaction between the two. Hypothesis 2 implies that the interaction term between ethnic minority and female should be positive and significant. As is evident from Figure 2 b) the estimate of the interaction effect is noisy, but the effect is substantial (11.2 percentage points) and significant ($p = 0.016$).

---

16 The fact that ethnicity is randomly assigned within occupations and gender between occupations implies that the effect of ethnicity is measured with more precision than the effect of gender. We account for this by using randomisation inference with the same randomisation procedure. Gender is clustered on the vacancy level and ethnicity is block randomised on the vacancy level. We impute constant additive effects and run 100,000 iterations. Alternative specifications using OLS regression with clustered standard errors generate similar results (appendix F1.)
While there is a large penalty for belonging to the ethnic minority group and a small (insignificant) penalty for being female within the majority group, these differences are not additive. Instead, the interaction term denotes that minority females receive a substantively higher call-back rate than we would expect if the ethnic and gender differences were purely additive.

As pointed out previously, heterogeneous effects across occupations can be decisive for the overall effects in correspondence experiments (Bursell, 2014; Midtbøen, 2016). Could the interaction effect between ethnicity and gender to some extent be an artefact of the sample’s composition of occupations? We explore this notion in two steps. In table 3, we re-weight call-back rates by occupational categories and examine the results given the
sample consisting of equally sized occupational categories.\textsuperscript{17} Although the relative differences are slightly smaller in the re-weighted sample, the same overall pattern is consistent. For female applicants, the re-weighted ratio is 1.18 compared to 1.22 in the original sample, while for male applicants the call-back ratio alters from 1.88 in the original sample to 1.76 in the re-weighted sample. The total ratio between majority and minority applicants is reduced from 1.52 to 1.44.

\begin{table}[h]
\centering
\caption{Call-backs re-weighted by occupational category}
\begin{tabular}{lccc}
\hline
 & \textit{Female} & \textit{Male} & \textit{Total} \\
\hline
Majority & 31.5 & 38.1 & 34.9 \\
Minority & 26.7 & 21.6 & 24.1 \\
Ratio & 1.18 & 1.76 & 1.44 \\
\hline
\end{tabular}
\end{table}

Secondly, we explore results in subsets of the sample based on the gender-stratification in the labour market reported in Table 4. We identify 112 jobs in female-dominated occupations (share of males < 20 pct.) and 90 jobs in male-dominated occupations (share of females < 20 pct.) while the rest is denoted as gender-balanced occupations. Female-dominated occupations in the sample jobs as nurses, pedagogues, social and health care assistants, social workers, receptionists and secretaries, and cleaning workers, while male-dominated occupations include building and construction workers, mechanics, warehouse workers, it-supporters, and drivers (appendix D).

The overall pattern in call-backs across ethnicity and gender is consistent in all three subsets of the data. In each subset, minority males received fewer call-backs than both majority males and majority and minority females. The relative difference between majority and minority applicants is most substantial in the male-dominated jobs, where minority males only received a call-back rate of 9.6 pct. compared to 36.5 pct. of the

\textsuperscript{17} Adjusting for occupational industry-dummies does not alter the estimates but slightly increases precision
majority males, implying a relative difference of 3.8. The smallest ratio of 1.19 is observed among female applicants in female dominated occupations where the ratio is 1.73 among male applicants. These results are aligned with Bursell’s (2014) correspondence experiment in the Swedish labour market.18

Table 4. Call-backs across gender-stratified and balanced occupations

<table>
<thead>
<tr>
<th>Gender balanced occupations</th>
<th>Male-dominated occupations</th>
<th>Female-dominated occupations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Majority</td>
<td>29</td>
<td>37.8</td>
</tr>
<tr>
<td>Minority</td>
<td>24</td>
<td>13.2</td>
</tr>
<tr>
<td>Ratio</td>
<td>1.2</td>
<td>1.54</td>
</tr>
<tr>
<td>N</td>
<td>100</td>
<td>98</td>
</tr>
</tbody>
</table>

In summary, the exploratory analysis shows that the effects do not seem to be an artefact of the composition of occupational categories in the sample. Moreover, the interaction between gender and ethnicity is present in both gender balanced, male-dominated and female-dominated occupations. The relative difference among majority and minority applicants is substantially larger in male-dominated jobs, suggesting that males both execute and become the target of ethnic discrimination more often than females.

Does SES confound the use of names as a signal for ethnicity?

To examine if the ethnic traits are confounded by notions of socioeconomic status, we randomly assigned the majority names from two groups: the first group consisted of the most popular Danish names, while the other group consisted of names from the bottom

18 We also examine the intersection between ethnicity and gender across sector and find the same pattern within both public sector jobs and private sector jobs (appendix G).
percentile of the SES-index. On this basis, we test if the difference between ethnic minority and majority candidates is confounded by the status of the majority applicants. We seek to estimate the differences in treatment effects of ethnicity conditional on the information about SES related to the majority applicants. Hence, the quantity of interest in this section is a difference in differences represented by the following expression:

$\left( E[\text{Call-back}|\text{Majority Popular}] - E[\text{Call-back}|\text{Minority competing against Majority Popular}] \right) - \left( E[\text{Call-back}|\text{Majority low-SES}] - E[\text{Call-back}|\text{Minority competing against Majority low-SES}] \right)$

This estimand captures the degree to which effects of SES are consequential for the relative difference in call-backs. If having a low-SES name reduce the chances of receiving a call-back, the estimand will be negative.

Table 5 shows call-back rates for competing pairs of majority and minority applicants. The immediate comparison of call-backs to majority applicants indicates that having a low-SES name reduces the chances of receiving a call-back by roughly four percentage points. However, the relative difference in call-backs between majority and minority applicants is actually larger when the majority candidate has a low-SES name (12.4 percentage points) compared to jobs where the majority candidate has a popular name (10.6 percentage points). Hence, the difference in differences is negative, which suggests that that ethnic discrimination is not altered by the status of the majority candidate. While the estimate is noisy, we can reject the third hypothesis that SES related to majority applicants’ names is a significant factor of the effect.

<table>
<thead>
<tr>
<th>Competing pairs when Majority candidate has a popular name</th>
<th>Competing pairs when Majority candidate has a low-SES name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>35.7</td>
</tr>
<tr>
<td>Minority</td>
<td>25.1</td>
</tr>
</tbody>
</table>
One concern is that our sample size may be too small to detect effects of SES but the fact that the difference-in-differences is actually negative supports the conclusion that providing information on SES among majority candidates does not affect levels of discrimination substantially. We also explore the intersections between gender, SES, and ethnicity but find no major differences from the overall pattern (appendix I).

In conclusion, the excludability assumption seems to hold: the differences between ethnic majority and ethnic minority candidates is not affected by the SES related to the majority applicants’ names. This bolsters the validity of the use of distinctive names to signal ethnicity.

7. Conclusion and discussion

This paper sheds new light on the disparate treatment of ethnic minorities in the labour market, a topic that has received substantial interest in the scholarly community as well as being the subject of intense discussions and public debate. The results show that when equally qualified applicants apply for a job, an applicant with a Middle Eastern-sounding name is significantly less likely to receive a call-back compared to an applicant with a Danish-sounding name. The overall result is similar to findings in previous correspondence studies conducted in a number of European countries. The difference is notable considering that the fictitious applicants used in this experiment were highly qualified for the jobs applied for, which should minimise the incentive to discriminate.
The findings demonstrate that discrimination varied substantially by gender, which serves as a call to future correspondence experiments to manipulate gender in order to estimate general measures of ethnic discrimination. The results undergird the notion that male ethnic minorities are especially exposed to discrimination – not only in specific occupational categories but across a broad spectrum of the labour market. Furthermore, this paper examined whether the socioeconomic status of names influences the chances of receiving a call-back. By actively choosing control names, we were able to examine if the ethnic gap decreased when the majority name belonged to a group of low socioeconomic status names. The results do not suggest that the majority applicants’ names confounded the effect of ethnicity, which bolsters the interpretation that the gap in call-backs can be attributed to ethnicity and not characteristics related to the use of distinct names.

Despite its innovations, the present study also has a number of limitations. First, this experiment proves discrimination against candidates with a given set of credentials in a non-representative sample of the Danish labour market. While we did apply for a broad array of jobs, it is impossible to draw a representative sample of job openings from the ever-changing job market. Hence, applicants applying for other types of jobs—and with different educational levels, experience or personal characteristics—might face different outcomes than found in the present study. It should also be noted that while we have studied discrimination in the first stage of the hiring process, discrimination could occur at the job interview or within a workplace in the wage-setting or promotion process (Pager & Western, 2012).

Secondly, despite the large difference in invitations for job interviews, it remains uncertain how discrimination affects the employment rate of the large group of Middle Eastern immigrants and descendants in Denmark more generally. Because the availability
of jobs and the strategy of the individual applicant plays a decisive role in getting a job, discrimination on average does not necessarily translate into gaps in employment rates (Heckman, 1998). Minorities might have prior knowledge about non-discriminating workplaces or find jobs outside of the formal hiring processes (Demireva, 2008), and thereby avoid discriminating employers. Nevertheless, the gap between majority and minority candidates found in this experiment is substantial and occurred across different occupational categories, from the largest Danish online job portal, and it therefore seems highly plausible that discrimination translates into overall employment disadvantages. Furthermore, the findings raise concerns about the presence of a meritocratic principle and suggest widespread violations of the law of equal treatment in the labour market.

Thirdly, a concern in correspondence experiments is that names are imprecise proxies of the groups they represent. In this case, the names used to signify gender and ethnic differences are distinct, and there is little reason to doubt that employers understand the traits. However, it could be argued that the names used to signify socioeconomic status do not match employers’ perceptions of low-SES names. The names were selected based on three socioeconomic factors (average crime, income and unemployment), and even though the names match a common notion of low-status names, we cannot be completely sure that these names were perceived as such. It should be mentioned, however, that the manipulation of SES served to bolster the validity of the ethnic trait—we were not interested in studying the effects of variations in SES in itself, in which case we would have manipulated more than just the name (Jackson, 2009). In addition, it should be mentioned that while we examined the importance of having a low-SES majority name, the application templates included information that signalled a middleclass lifestyle, which might crowd out the effect of low-SES names. Hence, we cannot ignore the
possibility that SES might be important if the applications were less informative or among less qualified applicants.

Finally, our findings leave open the important question of why we see these results. This experiment was primarily designed to examine if, and against whom, employers discriminated, and we cannot definitively establish the causal mechanisms underpinning the results. On the one hand, the fact that ethnic discrimination is gender-reliant supports the notion of Social Domination theory, which can be understood as a type of taste-based discrimination mechanism. On the other hand, this finding is seemingly consistent with the expectations of statistical discrimination models as well. If minority males perform, on average, worse than minority females do on some outcome-relevant characteristics, we might expect employers to treat the two groups differently. Female descendants with a Middle Eastern background outperform male descendants in a number of statistics (education, grades, wages, etc.) which employers might utilise in their evaluation of applicants (Statistics Denmark, 2017). In addition, employers’ perceptions of cultural distance might vary by the gender of the minority applicant. Cultural or value differences, for example views on gender equality, could be perceived as larger or more problematic when the minority candidate is male (Lancee, Soiné, Reino, & Veit, 2017). While the importance of hard skills (e.g. increased experience or reference letters) has been examined in previous studies (Arai et al., 2016; Kaas & Manger, 2012), little is known about the ways in which perceptions of cultural differences moderate ethnic discrimination. This could be examined in future research by manipulating cultural information, for example by signifying support for gender equality, democratic participation or religious affiliation.
Thus, an important task in future research is to increase the understanding of why ethnic minorities—especially males—are penalised. As this paper demonstrates, it is possible within the framework of correspondence experiments to study how different components affect behaviour among employers. We urge researchers on ethnic discrimination to replicate and extend work in this area by disentangling the effects of other components alone and in combination in order to contribute towards measuring and understanding ethnic discrimination.
## Literature


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


Mudde, C. (2013). Three decades of populist radical right parties in Western Europe: So


Appendix A

In this appendix, we describe our approach to selecting the names of the job candidates that were used in the experiment. The names were chosen from a dataset comprising the number of Danish citizens with a given name. The popular ethnic majority names were all among the top 10 most popular male and female names among those aged between 20 and 40. The ethnic minority names were identified from among the 20 most popular Middle Eastern names in use in Denmark, and all figure in a dictionary of Muslim first names that are in frequent use in Denmark (Meldgaard, 2005). Furthermore, the last names were chosen from among the most popular Danish-sounding and Middle Eastern-sounding names.

Finally, we identified low-SES names based on a population index. We constructed the SES index using three items: average crime rate, unemployment rate, and annual income. The crime rate is a measure of the proportion of the population with a given name who have been incarcerated within the last five years. The unemployment statistic is a measure of the proportion of people with a given name who receive unemployment benefits or have been unemployed for at least six months. Income is the average annual income earned by those bearing a given name. All items were scored from 1 to 8, where a higher score indicates lower status (with income scored in reverse order). The selected low-SES names all belong to individuals in the bottom decile of the SES index with an average age between 20 and 40.
<table>
<thead>
<tr>
<th>Majority names</th>
<th>Minority names</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low SES</strong></td>
<td><strong>High SES</strong></td>
</tr>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Jimmi</td>
<td>Bonnie</td>
</tr>
<tr>
<td>Sonny</td>
<td>Jennie</td>
</tr>
<tr>
<td>Ricky</td>
<td>Jennifer</td>
</tr>
<tr>
<td>Kenny</td>
<td>Belinda</td>
</tr>
<tr>
<td>Ronnie</td>
<td>Michelle</td>
</tr>
<tr>
<td>Nicky</td>
<td>Stella</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B
In this appendix, we test whether the chance of receiving a call-back is affected by the application process. We regress the outcome variable (call-backs) on three variables related to the application process: which application employers received first (the elapsed time between the two applications was 2–3 days), which of 16 unique applications was used, and which of two possible CVs was used. We calculate the heteroscedasticity-robust Wald statistic for the hypothesis that all the coefficients on the variables are zero (blue line). In order to calculate the related $p$-value, we use randomization inference to create a reference distribution of Wald statistics from 10,000 permutations. The results show no reason to reject the null hypothesis ($p$-value = .63).

Figure B. Call-backs across application, order and CV
Appendix C

In this appendix, we perform a balance check to assess whether the random assignment of gender and SES appear to be reliable. We test whether observed covariate imbalances are larger than expected from chance alone. The covariates include five dummy variables: size (more or less than 50 employees), sector (public or private), education (if tertiary education was required), language requirements (if the job required good language abilities) and customer contact (frequent customer contact). We calculate the heteroscedasticity-robust Wald statistic for the hypothesis that all the coefficients on the variables are zero (blue line). In order to calculate the related p-values, we use randomization inference to create a distribution of 100,000 Wald statistics. As Figure C1 and Figure C2 show, the results does not indicate a reason to reject the null hypothesis ($p = .43$ and $p = .92$).
Figure C2. Balance check for assignment of SES
Appendix D

In this appendix we show the descriptive statistics of the sample across the six occupational categories and the various occupations (23 in total). The table includes the number of applications that were sent for each occupation (N), the call-back rates for both majority and minority candidates, and the gender division in the specific occupation. Gender division is based on data from the Danish National Centre for Social Research (Larsen et al., 2016) and denoted Balanced if the share of males in a given occupation is between 20 and 80 per cent, Male if the share of females is < 20 per cent, and Female if the share of males is < 20 per cent.
<table>
<thead>
<tr>
<th>Occupational categories and occupations</th>
<th>N</th>
<th>Call-back Majority</th>
<th>Call-back Minority</th>
<th>Ratio</th>
<th>Private sector jobs</th>
<th>Gender division*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative support</td>
<td>66</td>
<td>.136</td>
<td>.106</td>
<td>1.29</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Accountancy</td>
<td>8</td>
<td>.125</td>
<td>.000</td>
<td></td>
<td>6</td>
<td>Balanced</td>
</tr>
<tr>
<td>Administrative assistant</td>
<td>20</td>
<td>.150</td>
<td>.150</td>
<td></td>
<td>16</td>
<td>Balanced</td>
</tr>
<tr>
<td>IT-support</td>
<td>15</td>
<td>.133</td>
<td>.133</td>
<td></td>
<td>13</td>
<td>Male</td>
</tr>
<tr>
<td>Receptionist &amp; secretary</td>
<td>23</td>
<td>.130</td>
<td>.087</td>
<td></td>
<td>21</td>
<td>Female</td>
</tr>
<tr>
<td><strong>Construction &amp; Extraction</strong></td>
<td>81</td>
<td>.358</td>
<td>.148</td>
<td>2.42</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Building and construction</td>
<td>4</td>
<td>.750</td>
<td>.000</td>
<td></td>
<td>4</td>
<td>Male</td>
</tr>
<tr>
<td>Carpenter</td>
<td>9</td>
<td>.333</td>
<td>.222</td>
<td></td>
<td>9</td>
<td>Male</td>
</tr>
<tr>
<td>Gardener</td>
<td>1</td>
<td>.000</td>
<td>.000</td>
<td></td>
<td>1</td>
<td>Balanced</td>
</tr>
<tr>
<td>Industrial production</td>
<td>7</td>
<td>.429</td>
<td>.143</td>
<td></td>
<td>6</td>
<td>Balanced</td>
</tr>
<tr>
<td>Mechanic</td>
<td>6</td>
<td>.500</td>
<td>.333</td>
<td></td>
<td>6</td>
<td>Male</td>
</tr>
<tr>
<td>Driver</td>
<td>25</td>
<td>.520</td>
<td>.200</td>
<td></td>
<td>25</td>
<td>Male</td>
</tr>
<tr>
<td>Warehouse</td>
<td>29</td>
<td>.138</td>
<td>.069</td>
<td></td>
<td>28</td>
<td>Male</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td>55</td>
<td>.382</td>
<td>.255</td>
<td>1.50</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Pedagoule</td>
<td>29</td>
<td>.310</td>
<td>.207</td>
<td></td>
<td>3</td>
<td>Female</td>
</tr>
<tr>
<td>Teacher</td>
<td>26</td>
<td>.462</td>
<td>.308</td>
<td></td>
<td>3</td>
<td>Balanced</td>
</tr>
<tr>
<td><strong>Health care</strong></td>
<td>60</td>
<td>.433</td>
<td>.400</td>
<td>1.08</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Nurse</td>
<td>14</td>
<td>.385</td>
<td>.462</td>
<td></td>
<td>3</td>
<td>Female</td>
</tr>
<tr>
<td>Rehabilitation assistant</td>
<td>17</td>
<td>.294</td>
<td>.353</td>
<td></td>
<td>1</td>
<td>Balanced</td>
</tr>
<tr>
<td>Social and health care assistant</td>
<td>22</td>
<td>.455</td>
<td>.364</td>
<td></td>
<td>1</td>
<td>Female</td>
</tr>
<tr>
<td>Social worker</td>
<td>8</td>
<td>.750</td>
<td>.500</td>
<td></td>
<td>1</td>
<td>Female</td>
</tr>
<tr>
<td><strong>Retail</strong></td>
<td>98</td>
<td>.306</td>
<td>.173</td>
<td>1.76</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>Cleaning</td>
<td>17</td>
<td>.294</td>
<td>.059</td>
<td></td>
<td>14</td>
<td>Female</td>
</tr>
<tr>
<td>Restaurant &amp; Hotel</td>
<td>19</td>
<td>.263</td>
<td>.158</td>
<td></td>
<td>19</td>
<td>Balanced</td>
</tr>
<tr>
<td>Retail and service</td>
<td>62</td>
<td>.323</td>
<td>.210</td>
<td></td>
<td>58</td>
<td>Balanced</td>
</tr>
<tr>
<td><strong>Sales &amp; related</strong></td>
<td>40</td>
<td>.475</td>
<td>.350</td>
<td>1.36</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Communication and marketing</td>
<td>6</td>
<td>0.333</td>
<td>0.333</td>
<td></td>
<td>6</td>
<td>Balanced</td>
</tr>
<tr>
<td>Salesman</td>
<td>12</td>
<td>0.250</td>
<td>0.083</td>
<td></td>
<td>11</td>
<td>Balanced</td>
</tr>
<tr>
<td>Telemarketing</td>
<td>22</td>
<td>0.636</td>
<td>0.500</td>
<td></td>
<td>22</td>
<td>Balanced</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>400</td>
<td>0.335</td>
<td>0.220</td>
<td></td>
<td>278</td>
<td></td>
</tr>
</tbody>
</table>
Appendix E

In this appendix, we visualise the approximate permutation tests for the effect of the ethnic trait among all candidates, female candidates, and male candidates, respectively. We used the R package randomizr to perform these tests. We also visualise the approximate permutation test for the interaction between gender and ethnicity.

The fact that each unit received both treatment and control increases precision. Rather than comparing call-backs across a heterogeneous collection of employers, we increase statistical power by controlling for employer-level heterogeneity. We assess statistical uncertainty around each estimate by running 100,000 permutations with block random assignment of ethnicity at the job level under the sharp null assumption of no effects for all units. This corresponds to an OLS regression with fixed effects.

While the ethnic indicator variable is manipulated within jobs, gender is manipulated across jobs. Hence, the interaction model between ethnicity and gender is analysed using 100,000 permutations under constant, additive treatment effects with block random assignment of ethnicity and cluster random assignment of gender at the job level. Hence, Figure E1 presents randomization inference for the ethnic difference among A) all applicants, B) female applicants and C) male applicants. Finally, D) presents the interaction between minority status and gender.
Figure E1. Hypothesis tests

A) SE: .024; p = 0

B) SE: .032; p = .099

C) SE: .036; p = 0

D) SE: .047; p = .016
Appendix F

In this appendix we show three alternative specifications of the interaction models. Column (1) reports the covariate-unadjusted model, while column (2) reports the covariate-adjusted model. The fact that ethnicity is randomly assigned within occupations while gender is assigned between occupations implies that the effect of ethnicity is measured with more precision than the effect of gender. We account for this by using randomization inference with the same randomization scheme – gender is clustered at the vacancy level and ethnicity block randomized at the vacancy level – with 100,000 iterations and the effect imputed for each iteration to obtain standard errors. Finally, to compare results with a conventional OLS regression model, column (3) reports the covariate-unadjusted model with vacancy fixed effects and cluster corrected standard errors.

The rationale behind including covariates is to reduce disturbance variability, but the inclusion of covariates does little to improve the precision of our estimated treatment effects (in fact, SE increases slightly in the adjusted model). The covariates include five dummy variables: size (more or less than 50 employees), sector (public or private), education (if tertiary education was required), language requirements (if the job required good language abilities) and customer contact (frequent customer contact).

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minority</td>
<td>-.169</td>
<td>-.169</td>
<td>-.169</td>
</tr>
<tr>
<td></td>
<td>(.033)</td>
<td>(.034)</td>
<td>(.034)</td>
</tr>
<tr>
<td>Female</td>
<td>-.056</td>
<td>-.068</td>
<td>-.056</td>
</tr>
<tr>
<td></td>
<td>(.044)</td>
<td>(.047)</td>
<td>(.047)</td>
</tr>
<tr>
<td>Minority * Female</td>
<td>.112</td>
<td>.113</td>
<td>.112</td>
</tr>
<tr>
<td></td>
<td>(.047)</td>
<td>(.047)</td>
<td>(.046)</td>
</tr>
<tr>
<td>Covariate adjustment</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

N = 400

Standard errors are in parentheses.

19 The covariates are mean centered and interacted with the treatment-indicators.
**Interaction between minority status and SES of the majority applicant**

Column (1) reports the covariate-unadjusted model, while column (2) reports the covariate-adjusted model. We estimate uncertainty by using randomization inference with the same randomization scheme – SES is clustered at the vacancy level and ethnicity block randomized at the vacancy level – with 100,000 iterations and the effect imputed for each iteration to obtain standard errors.

Finally, to compare results with a conventional OLS regression model, column (3) in reports results from a regression of a call-back dummy on the Minority-indicator, the SES-indicator, and the interaction between the two. Standard errors are cluster corrected.

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minority</td>
<td>-.124</td>
<td>-.121</td>
<td>-.124</td>
</tr>
<tr>
<td></td>
<td>(.033)</td>
<td>(.043)</td>
<td>(.033)</td>
</tr>
<tr>
<td>Popular name</td>
<td>.043</td>
<td>.05</td>
<td>.043</td>
</tr>
<tr>
<td></td>
<td>(.047)</td>
<td>(.044)</td>
<td>(.047)</td>
</tr>
<tr>
<td>Minority * Popular name</td>
<td>-.018</td>
<td>-.012</td>
<td>-.018</td>
</tr>
<tr>
<td></td>
<td>(.048)</td>
<td>(.062)</td>
<td>(.047)</td>
</tr>
<tr>
<td>Covariate adjustment</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>N = 400</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are in parentheses.
Appendix G

In this appendix, we report the call-back rates across the intersection between sector, ethnicity and gender.

Table G1. Call-back rate across sector, ethnicity and gender

<table>
<thead>
<tr>
<th></th>
<th>Private sector</th>
<th></th>
<th>Public sector</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>All</td>
<td>Female</td>
</tr>
<tr>
<td>Majority</td>
<td>.277</td>
<td>.331</td>
<td>.306</td>
<td>.365</td>
</tr>
<tr>
<td>Minority</td>
<td>.231</td>
<td>.135</td>
<td>.179</td>
<td>.33</td>
</tr>
<tr>
<td>Ratio</td>
<td>1.2</td>
<td>2.5</td>
<td>1.7</td>
<td>1.1</td>
</tr>
<tr>
<td>N</td>
<td>278</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix H

To enhance transparency, the hypotheses were preregistered before any data was collected. The preregistration is found at http://egap.org/ with ID 20150930AA. The preregistration document contains a very brief introduction to the study, the design, the three hypotheses and the outcome of interest. It should be noted that the preregistration lacks important information such as: sampling frame, attrition criteria, a specific plan with technical details for the data analysis and explicit power calculations. This is all information that would of course have been appropriate to include.
Appendix I

This appendix reports call-backs across gender, ethnicity and SES of the majority candidates and the difference-in-differences.

Table I. Call-backs across ethnicity, SES and gender

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Call-backs when Majority</td>
<td>Call-backs when Majority</td>
</tr>
<tr>
<td></td>
<td>candidate has a popular name</td>
<td>candidate has a low-SES name</td>
</tr>
<tr>
<td>Majority</td>
<td>.409</td>
<td>.309</td>
</tr>
<tr>
<td>Minority</td>
<td>.236</td>
<td>.144</td>
</tr>
<tr>
<td>Difference</td>
<td>.173</td>
<td>.165</td>
</tr>
<tr>
<td>N</td>
<td>110</td>
<td>97</td>
</tr>
<tr>
<td>Difference</td>
<td>0.008 [0.007]</td>
<td></td>
</tr>
</tbody>
</table>

Standard errors reported in parentheses are clustered at the vacancy level
Appendix J

In this appendix we report the distribution of the number of jobs applied for and call-backs across the minority candidates’ names to examine whether the religious connotations of certain minority names affect the likelihood of receiving a call-back. An F-test for the hypothesis that call-backs to candidates across names are equivalent yields p-values of .81 and .13.

**J1. Call-backs male candidates**

<table>
<thead>
<tr>
<th>Name</th>
<th>N</th>
<th>Call-back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdul</td>
<td>34</td>
<td>.205</td>
</tr>
<tr>
<td>Ahmad</td>
<td>22</td>
<td>.136</td>
</tr>
<tr>
<td>Ali</td>
<td>23</td>
<td>.173</td>
</tr>
<tr>
<td>Hassan</td>
<td>22</td>
<td>.136</td>
</tr>
<tr>
<td>Ibrahim</td>
<td>29</td>
<td>.275</td>
</tr>
<tr>
<td>Mohammad</td>
<td>29</td>
<td>.172</td>
</tr>
<tr>
<td>Mustafa</td>
<td>26</td>
<td>.269</td>
</tr>
<tr>
<td>Omar</td>
<td>22</td>
<td>.136</td>
</tr>
</tbody>
</table>

H0: name effects are zero (p-value) 0.813

**J2. Call-backs female candidates**

<table>
<thead>
<tr>
<th>Name</th>
<th>N</th>
<th>Call-back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aisha</td>
<td>22</td>
<td>.318</td>
</tr>
<tr>
<td>Amal</td>
<td>19</td>
<td>.263</td>
</tr>
<tr>
<td>Amira</td>
<td>22</td>
<td>.136</td>
</tr>
<tr>
<td>Fatima</td>
<td>27</td>
<td>.296</td>
</tr>
<tr>
<td>Fatma</td>
<td>29</td>
<td>.206</td>
</tr>
<tr>
<td>Hatice</td>
<td>25</td>
<td>.24</td>
</tr>
<tr>
<td>Iman</td>
<td>25</td>
<td>.44</td>
</tr>
<tr>
<td>Zainab</td>
<td>24</td>
<td>.08</td>
</tr>
</tbody>
</table>

H0: name effects are zero (p-value) 0.134