WORKING PAPER

THE STATE OF HIRING DISCRIMINATION: A META-ANALYSIS OF (ALMOST) ALL RECENT CORRESPONDENCE EXPERIMENTS

Louis Lippens
Siel Vermeiren
Stijn Baert

December 2021
2021/1035

GHENT UNIVERSITY

Department of Economics

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

The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments*

Louis Lippens,[i] Siel Vermeiren,[ii] and Stijn Baert[iii]

Abstract

Notwithstanding the improved integration of various minority groups in the workforce, unequal treatment in hiring still hinders many individuals' access to the labour market. To tackle this inaccessibility, it is essential to know which and to what extent minority groups face hiring discrimination. This meta-analysis synthesises a quasi-exhaustive register of correspondence experiments on hiring discrimination published between 2005 and 2020. Using a random-effects model, we computed pooled discrimination ratios concerning ten discrimination grounds upon which unequal treatment in hiring is forbidden under United States federal or state law. We find that hiring discrimination against candidates with disabilities, older candidates, and less physically attractive candidates is at least equally severe as the unequal treatment of candidates with salient racial or ethnic characteristics. Remarkably, hiring discrimination against older applicants is even higher in Europe than in the United States. Furthermore, unequal treatment in hiring based on sexual orientation seems to be prompted mainly by signalling activism rather than same-sex orientation in itself. Last, aside from a significant decrease in ethnic and racial hiring discrimination in Europe, we find no structural evidence of recent temporal changes in hiring discrimination based on the various other grounds within the scope of this review.

Keywords: hiring discrimination, unequal treatment, meta-analysis, correspondence experiment, audit study

JEL Classification: J71, J23, J14, J15, J16

---

* Acknowledgements. We are grateful to the authors of a handful of the correspondence experiments included in this review for providing us with missing data to supplement our dataset and with feedback on a previous version of this manuscript. We are also thankful to Brecht Neyt for his helpful comments and suggestions.

[i] Corresponding author. Ghent University, Sint-Pietersplein 6, 9000 Ghent, Belgium, and Vrije Universiteit Brussel. Louis.Lippens@UGent.be. ORCID: 0000-0001-7840-2753.

[ii] Ghent University.

[iii] Ghent University, University of Antwerp, Université catholique de Louvain, IZA Institute of Labor Economics, Global Labor Organization, and IMISCOE International Migration Research Network. ORCID: 0000-0002-1660-5165.

1. Introduction

Although the workforce has become increasingly diverse—improving the integration of female, migrant, and older workers, amongst other groups—many individuals belonging to various minority groups still face considerable discrimination in the labour market (Organisation for Economic Co-operation and Development [OECD], 2020a). In part because of their decreased chances for labour market access, these individuals are at elevated risk of long-term unemployment and labour market inactivity (OECD, 2020a, 2020b). This underutilisation of talent could result in needless economic costs for society (OECD, 2020a; Pager, 2016). For policymakers, it is vital to know which (minority) groups are confronted with hiring discrimination and understand the severity of this labour market inaccessibility. In this way, targeted diversity policies (e.g. outreach campaigns targeting under-represented groups) can be implemented to help those who require said policies the most (OECD, 2020a).

Research on labour market discrimination has long focused on the non-experimental decomposition approach to measure discrimination (Blinder, 1973; Neumark, 2018; Oaxaca, 1973). This approach has historically involved isolating the impact of discrimination on wages via regression analyses (Borjas, 2020). Variance that could not be explained by the differences in human capital between the groups of interest (e.g. blacks and whites) was consequently attributed to discrimination. However, it is difficult to capture the true amount of variance explained by human capital under this approach, primarily due to omitted variable bias (Altonji & Blank, 1999; Borjas, 2020).[1] The decomposition method thus sketches an incomplete picture of discrimination (Borjas, 2020; Gaddis, 2018).

To overcome the limitation of the decomposition approach, researchers had begun to use audit studies as an alternative experimental method to measure the incidence of labour market discrimination (Gaddis, 2018). At first, this was mainly done by sending out pairs of real applicants (i.e. actors) who differed in terms of visible characteristics based on which unequal treatment is forbidden (e.g. skin colour) to interview for the same job. Differences in job offers were subsequently interpreted in terms of discrimination. In the early 2000s,

[1] Examples of omitted variables include (but are not limited to) unobserved supply-side factors such as personal motivation or choices as well as possessing an extensive professional network.

however, Bertrand and Mullainathan (2004) steered the research area of labour market discrimination in a different direction: in-person audits replaced correspondence audits as the standard for measuring hiring discrimination (Gaddis, 2018). Rather than sending out actors as applicants, these correspondence experiments consisted of mailing written applications from fictitious job seekers in response to real job postings. By randomly assigning individual characteristics based on which hiring selection is forbidden, the effect of these characteristics on employers' reactions could be given causal interpretations. Compared with in-person audits, the perceived differences between applicants produced by minute differences in their behaviour during the interview were nullified. Moreover, the application process was less resource-intensive. Because of its high employability and the causal interpretation that underpins its results, the correspondence testing method is still widely used today (Baert, 2018; Neumark, 2018).

In recent years, a considerable number of scholars have reviewed and synthesised parts of the hiring discrimination literature, focusing on the correspondence testing method. We know of eleven contemporary meta-studies in this area: Baert (2018), Bartkoski et al. (2018), Bertrand and Duflo (2016), Flage (2020), Heath and Di Stasio (2019), Neumark (2018), Quillian and Midtbøen (2021), Quillian et al. (2017, 2019), Rich (2014), Thijssen et al. (2021b), and Zschirnt and Ruedin (2016). Several of these studies have provided insightful observations related to theory or policy but, for understandable reasons, have only focused on certain grounds of discrimination (predominantly race and ethnicity) while neglecting others (e.g. Bartkoski et al., 2018; Flage, 2020; Quillian et al., 2017). Other reviews taking a broader view of hiring discrimination have brought forth equally interesting insights but are rather narrative in nature, instead of providing a systematic account of the existing literature (e.g. Bertrand & Duflo, 2016; Neumark, 2018; Rich, 2014). Baert (2018) attempted to counter these limitations by (i) adopting a broad view on hiring discrimination that considered all grounds based on which unequal treatment is forbidden under United States law and (ii) providing a quasi-exhaustive register of correspondence experiments conducted since Bertrand and Mullainathan's (2004) seminal study. However, the main limitation of Baert's (2018) work was the absence of a meta-analytical component that synthesised the results of the included studies.

The current study documents and synthesises the most extensive register of correspondence experiments on hiring discrimination to date. Building on Baert’s (2018) work, we compiled a comprehensive catalogue of correspondence experiments from 2005 to 2020, simultaneously providing an overview of the hiatuses that still exist in the current literature on hiring discrimination. Specifically, we brought together 306 correspondence experiments (i.e. units of observation) originating from 169 studies. Altogether, these experiments comprised over 900,000 fictitious applications in response to actual job vacancies. More importantly, we meta-analytically quantified hiring discrimination regarding ten discrimination grounds upon which unequal treatment is forbidden under United States federal or state law, including (i) race and national origin, (ii) gender and motherhood status, (iii) religion, (iv) disability, (v) age, (vi) military service or affiliation, (vii) wealth, (viii) marital status, (ix) sexual orientation, and (x) physical appearance.

From a policy perspective, our synthesis offers policymakers a contextualised understanding of the severity of hiring discrimination concerning the various grounds within the scope of this review. First and foremost, our standardised meta-analytical approach enables comparisons of levels of hiring discrimination amongst discrimination grounds and specific minority groups. Moreover, our quantitative approach identifies sub-group differences, providing a granular perspective of our findings: for each discrimination ground, we were able to evaluate whether (i) levels of hiring discrimination are related to how call-backs are reported and measured, (ii) persistent regional differences in hiring discrimination exist, and (iii) unequal treatment in hiring has changed over time.

2. Data and methods

In this section, we elaborate on (i) the scope of our meta-analysis; (ii) how we identified and selected studies; (iii) which variables we collected from these studies, as well as how we classified some of them into broader categories to identify differences across categories; and (iv) the details of the meta-analytical methods we used to analyse the resulting data. In this process, we paid special attention to Havránek et al.’s (2020) reporting guidelines for meta-analyses in economics—we refer the reader to Table A1 (in the appendix) for the corresponding checklist.

2.1. Scope

We used various eligibility criteria based on the Population, Intervention, Comparison, Outcome (PICO) framework to delineate our review (Richardson et al., 1995).[2] Table 1 provides an overview of these criteria. We limited our review to correspondence studies in which unequal treatment in hiring was assessed between fictitious applicants belonging to minority groups and their majority counterparts. More specifically, we considered studies written in English that were first published as a discussion paper, pre-print, or journal article between 2005 (the year after Bertrand and Mullainathan's seminal 2004 correspondence study) and 2020 (the most recent full calendar year at the time this study was conducted).[3]

< Table 1 about here >

Similar to the delineation of the discrimination grounds in Baert's (2018) correspondence experiment register, we limited our scope to the forms of hiring discrimination prohibited under United States federal or state law. We thus took into account the following discrimination grounds: (i) race and national origin, (ii) gender and motherhood status, (iii) religion, (iv) disability, (v) age, (vi) military service or affiliation, (vii) wealth, (viii) genetic information, (ix) citizenship status, (x) marital status, (xi) sexual orientation, (xii) political affiliation, (xiii) union affiliation, and (xiv) physical appearance. In our final analysis, we retained only ten of these grounds because, similar to Baert (2018), (i) no (new) correspondence experiments related to genetic information or citizenship status were identified in the search process; and (ii) we found only one experiment related to political orientation and one experiment related to union affiliation, based on which we could not conduct meta-analyses.

[2] We extended the PICO framework to be more specific in the delineation of the scope of our review. Specifically, we also considered 'study type', 'context', and 'timing' and excluded 'intervention' because it was not relevant to our search query.

[3] The specific year allocated to a given study was based on the year the study was initially published. For example, Larsen and Di Stasio (2021) was first published online as a pre-print in 2019 before appearing officially in 2021 in the *Journal of Ethnic and Migration Studies*. This allocation strategy coincides with the approach of Baert (2018).

2.2. Study selection

We used multiple sources to identify, screen, and select eligible studies for our meta-analysis. Figure 1 depicts a structured overview of this process. First, we identified potentially eligible correspondence studies. On the one hand, we sourced studies included in Baert's (2018) register of correspondence experiments, which resulted from an elaborate systematic search for correspondence experiments on hiring discrimination. On the other hand, we performed a systematic search on the Web of Science and Google Scholar databases in spring 2021. Our search used the keywords 'correspondence experiment', 'correspondence study', 'fictitious resume', 'fictitious cv', 'fictitious application', and 'field experiment' in combination with the keyword 'discrimination'. In general, we confined our search to studies published in the period 2005 to 2020. To extend this systematic search, we also performed a cited reference search with the references from Baert's (2018) work as the input of our queries.

< Figure 1 about here >

Next, we appraised the studies that had not already been identified by Baert (2018). In total, we evaluated the titles and abstracts of 933 studies against our eligibility criteria (see section 2.1). After an initial screening of the titles and abstracts, we reviewed the full text of the remaining 137 articles. The risk of reviewer bias was reduced by having two researchers independently review the selected articles. After this review, 79 studies were identified that fully matched the criteria. There were four reasons for excluding certain studies after appraising their full text: (i) unequal treatment based on the discrimination ground in the scope of the study was not forbidden under United States law (*N* = 27, 46.55% of the total number of excluded full texts); (ii) the correspondence experiment was entirely based on data used in a previously published (and already included) study (*N* = 20, 34.48%); (iii) the study did not use the correspondence testing method (e.g. in-person audit; *N* = 10, 17.24%); or (iv) the study was related to housing discrimination rather than hiring discrimination (*N* = 1, 1.72%).

We retained a total of 169 studies, of which 90 were already included in Baert's (2018) study, resulting in 306 units of observation. There are more units of observation than studies because of our definition of a 'unit of observation', which we defined as a unique correspondence experiment based on the related (i) discrimination ground, (ii) treatment group, (iii) control group, and (iv) region where the test was performed.[4]

2.3. Data collection

We captured a multitude of variables for each correspondence experiment. First, we registered the basic information of the studies, including the authors' names and the year the article was officially published. In addition to the latter, we also recorded (i) the year the study was initially published (e.g. as a pre-print or early-access article), which was the year we used when evaluating the article against our eligibility criteria, and (ii) the year the field experiment ended. Second, we documented where the research took place, including the country and (sub-)region. The latter was based on the M49 Standard for geographic regions of the United Nations (2021; see Table A2 for a tabulated overview). Third, we registered the (experimental) treatment group and the control group of the field experiment. The specific treatment groups identified in the included studies were classified into broader groups to facilitate further analyses. Because no common international framework of ethnic and racial minority groups exists, the classification of these groups consisted of a proprietary framework based on how various governmental bodies of OECD member countries collect and categorise diversity data (Balestra & Fleischer, 2018; European Commission, 2021; Morning, 2008).[5] This classification can be found in Table A3.

Fourth, we documented data related to the outcomes of the correspondence experiments. We captured the overall treatment effect (averaged across sub-groups at the experiment level) of the results in the original studies.[6] We also recorded the classification of the outcome variable (i.e. call-back). If a call-back consisted of an invitation to a job

---

[4] For example, Di Stasio et al. (2021) considered hiring discrimination against Muslims in Germany, Norway, Spain, the Netherlands, and the United Kingdom. To allow for sub-group analyses on the basis of region (see section 2.4), this study was subdivided into multiple units of observation stemming from the same study.

[5] For example, in their correspondence test, Jacquement and Yannelis (2012) assigned African American names to the minority group, while Gaddis (2015) used black-sounding names. In the United States and the United Kingdom, these origins are both classified as 'African (American)' or 'black'. Therefore, we created the category 'African/African American/Black' as an umbrella term for similar treatment groups.

[6] We adhered to the reporting in the original studies when we recorded the overall treatment effect (e.g. a positive, significant effect of the applicant's gender on the employer's response was recorded as such).

interview (or any broadly defined positive response of the employer, such as a request for additional information), we labelled it *sensu stricto* (or *sensu lato*).

Most importantly, we registered the number of observations and the number of positive call-backs in both the treatment and control groups. The accuracy of the variables related to the treatment effect was independently assessed and verified by at least two researchers. Outcome measure data were missing for 32 units of observation (9.82% of the total number of units, *N* = 326). After contacting the corresponding authors of the respective studies to retrieve these data, 12 cases could be completed (37.50% of cases with missing data), meaning that we had no data for the remaining 20 units of observation.[7] These units were excluded from the meta-analysis, resulting in 306 valid units. Reporting bias, which could (partly) originate from missing data, was formally evaluated in our robustness analyses (see section 2.4.3).

From these data, we derived a standardised 'discrimination ratio'. The specification of this ratio is shown in Equation 1. The discrimination ratio is a risk ratio (or relative risk) equal to the division of two proportions: (i) the proportion of positive call-backs in the treatment group (a_k) relative to the total number of observations in that group ($n_{k\,treat}$), and (ii) the proportion of positive call-backs in the control group (c_k) relative to the total number of observations in that group ($n_{k\,control}$). Because the discrimination ratio can be interpreted in terms of relative change, a ratio of 0.75 (for example) indicates a 25% reduction in positive call-backs of the (fictitious) applicants of the minority group vis-à-vis the applicants of the control group, aggregated at the level of the correspondence experiment. Since our estimation strategy assumed that the included discrimination ratios follow a normal distribution, we log-transformed these ratios before pooling them in our meta-analysis (see section 2.4.1). This approach ensured that opposite, same-sized effects were equidistant. Equation 2 illustrates the calculation of the standard error of the log-transformed discrimination ratios (see Harrer et al., 2021).

[7] These missing data were linked to ten studies: Beam et al. (2020), Carlsson and Eriksson (2019), Darolia et al. (2016), Drydakis (2017), Guul et al. (2019), Patacchini et al. (2015), Stone and Wright (2013), Thijssen et al. (2021a), Thomas (2018), and Yemane and Fernández-Reino (2021). Our intention is to incorporate the results of these studies in later iterations of this manuscript, given that these data will have become available to us. Based on the results of the robustness analyses, we do not expect that the integration of these data in our current dataset will have a major impact on the meta-analytic results in these iterations (see section 2.4.3 and Table A6).

$$DR_k = \frac{a_k/n_{k\ treat}}{c_k/n_{k\ control}} \tag{1}$$

$$SE_{\ln DR} = \sqrt{\frac{1}{a_k} + \frac{1}{c_k} - \frac{1}{n_{k\ treat}} - \frac{1}{n_{k\ control}}} \tag{2}$$

2.4. Analyses

Our synthesis was based on the results of the correspondence experiments identified and selected in the previous steps. Our goals were (i) to quantify and compare the level of hiring discrimination for each of the various discrimination grounds and treatment groups in the scope of our analysis and (ii) to identify sub-group differences based on (a) the definition of the call-back variable, (b) the region where the correspondence experiment took place, and (c) the time period related to the research. We used R (version 4.1.0) for our analyses and relied on the {meta} package for most of our calculations (e.g. estimating the pooled ratios, performing sub-group analyses, and detecting reporting bias; Balduzzi et al., 2019). We also used the {dmetar} package to identify influential cases, the {metasens} package to perform 'limit' meta-analyses, and the {metafor} package to examine the statistical (in)dependence of the sampled discrimination ratios (Harrer et al., 2019; Schwarzer et al., 2020; Viechtbauer, 2010).

2.4.1. Pooled discrimination ratios

To quantify the level of hiring discrimination across the various discrimination grounds, we used a random-effects model to pool the discrimination ratios of the included studies by discrimination ground and treatment group.[8] We opted for this model because it starts from the premise that the true level of discrimination varies across studies. We assumed that there was at least some variation in these levels caused by (subtle) differences in (i) the definition and conceptualisation of the treatment and control groups, (ii) the measurement of the responses (or call-backs), and (iii) the overall correspondence experimental process. We used Knapp–Hartung adjustments to calculate the confidence intervals around the

[8] To clarify, we did not perform one large meta-analysis comprising all correspondence tests within the scope of this review, which would generate very little insight into the drivers of the underlying discrimination.

pooled discrimination ratio estimates (Knapp & Hartung, 2003). This method assumes a *t*-distribution of the pooled effect rather than a normal distribution, which reduces the chance of obtaining false-positive results (Langan et al., 2019).[9]

We used the commonly reported Mantel–Haenszel method for binary outcome data to calculate the weights of the studies (*w*) in the reported pooled discrimination ratios—the formula is shown in Equation 3 (for more details, see Borenstein et al., 2009; Mantel & Haenszel, 1959). This method takes into account the number of cases in the treatment and control groups wherein the call-back was positive (*a* and *c*, respectively), as well as the number of cases in the treatment and control groups wherein the call-back was negative or absent (*b* and *d*, respectively; Mantel & Haenszel, 1959). This approach inherently attaches more importance to studies with larger sample sizes or higher numbers of positive call-backs. To generate more balanced weights, the weights were adjusted for between-study variance (τ^2) to decrease the overemphasis (or underemphasis) on studies with a relatively large (or small) sample size (see also section 2.4.2; Borenstein et al., 2009). Subsequently, these variance-adjusted weights ($w*$) were plugged into the general specification of the random-effects model, as illustrated in Equation 4. Here, $\widehat{DR}$ is the pooled discrimination ratio, DR_k represents the observed discrimination ratio of the individual correspondence experiments, ζ is the error related to the overarching distribution of true discrimination ratios, and ε symbolises the sampling error (Borenstein et al., 2009; Harrer et al., 2021).

$$w_k = \frac{(a_k + b_k) * c_k}{a_k + b_k + c_k + d_k} \tag{3}$$

$$\widehat{DR} = \frac{\sum_{k=1}^{K}(DR_k + \zeta_k + \varepsilon_k) * w_k^*}{\sum_{k=1}^{K} w_k^*} \tag{4}$$

2.4.2. Heterogeneity analyses

To meaningfully interpret the pooled discrimination ratios by discrimination ground and treatment group and to identify sub-group differences in discrimination levels, we quantified and examined variability in statistical and design-related heterogeneity. First, we assessed statistical heterogeneity by calculating two statistics that captured the variability

[9] The Knapp–Hartung adjustments therefore produce more conservative (i.e. wider) confidence interval estimates than when these adjustments would not be applied.

in the true discrimination ratios underlying the data (Rücker et al., 2008). More specifically, we first calculated I^2 estimates, which indicate the between-study variability in the true discrimination ratios not caused by sampling error (Harrer et al., 2021; Higgins & Thompson, 2002). The I^2 statistic compares the studies' discrimination ratios to the pooled ratio, weighted by the inverse of the variance of the respective studies, taking into account the total number of studies. Because of the latter, this statistic is insensitive to (substantial) changes in the number of studies included in the analysis (Cochran, 1954; Harrer et al., 2021; Hoaglin, 2016).[10] However, a notable drawback of the I^2 statistic is that it increases as the sampling error decreases and tends towards one as the sample sizes of the studies in the meta-analysis increase (Harrer et al., 2021; Rücker et al., 2008). To overcome this limitation, and as recommended by IntHout et al. (2014) and Veroniki et al. (2016), we also computed and reported 95% prediction intervals ($\widehat{DR}^* \; CI_{95\%}$), which rely on the standard error of the pooled effect and the between-study variance estimate τ^2. The latter was calculated using the Paule–Mandel method (Paule & Mandel, 1982). These intervals provide a range of ratios within which the discrimination ratios of future studies will fall in approximately 95% of the cases and over repeated sampling.

Second, we evaluated design-related heterogeneity (i.e. heterogeneity due to differing designs across studies) by performing sub-group analyses based on the following sub-groups: (i) treatment (group) classification, (ii) call-back classification, (iii) the region in which the correspondence experiment took place, and (iv) the period during which the experiment ended. This approach contributed to (partly) explaining the statistical heterogeneity estimated in the previous step. Following Schwarzer et al.'s (2015) guidelines, we only performed sub-group analyses of the groups of studies for which the total number of included studies was equal to or greater than ten. The call-back classification contained two levels: (a) *sensu stricto* (i.e. an invitation for an interview) and (b) *sensu lato* (i.e. a positive reaction from the employer, such as a request for additional information; see section 2.3). The geographical segmentation comprised four regions: (a) the Americas, (b) Europe, (c)

[10] Higgins and Thompson's (2002) guidelines state that values around 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively.

Asia, and (d) other (i.e. Oceania and Africa combined). Finally, in the last sub-group analysis, the sub-periods spanned (a) 2002 to 2010 and (b) 2011 to 2020.[11] To formally assess sub-group heterogeneity, we performed an omnibus *Q*-test to examine the overall difference between the sub-group levels—its null hypothesis assumes that there is no difference (Cochran, 1954).

2.4.3. Robustness analyses

Finally, to assess the robustness of our results, we measured and controlled for outliers (i.e. influential cases that substantially affected the pooled discrimination ratio), reporting bias (i.e. the under- or over-reporting of research findings due to the nature and direction of the research results), and potential statistical dependence between the sampled discrimination ratios (Borenstein et al., 2009; Higgins et al., 2019). First, we identified outliers by looking at studies with extremely small and large discrimination ratios. We defined said ratios as those for which the upper (lower) bound of the 95% confidence interval was lower (higher) than the lower (upper) bound of the confidence interval of the pooled discrimination ratio (Harrer et al., 2021). To clarify, this means that the ratios of these influential cases were so extreme that they significantly differed from the pooled ratio (at the 5% significance level). Eventually, we recalculated the pooled discrimination ratios, excluding these outliers.

Next, we employed various techniques to control for reporting bias because we did not explicitly account for all reporting biases (e.g. publication bias or language bias).[12] To this end, we used Peter et al.'s (2006) binary-effects adaptation of Egger's regression test to calculate a 'bias statistic' for funnel plot asymmetry, which compares the discrimination ratios of the respective studies against their standard errors—its null hypothesis assumes that there is no asymmetry. Furthermore, we performed 'limit' meta-analyses in which we

[11] There is a discrepancy between the temporal period of this review (2005–2020) and the timeframe used in the sub-group analyses. The latter is based on the year in which the correspondence experiment ended. The rationale for this is that this time variable more accurately represents the timing of the experiment (vis-à-vis the year the research was published).

[12] On the one hand, publication bias could arise because correspondence experiments with statistically significant results might be more appealing to publish than experiments that produce statistically insignificant results. On the other hand, language bias could be an issue, as only studies in English were included in the review.

allowed for interactions between the observed effects, on the one hand, and the standard error of the pooled effect and the between-study variance, on the other hand (Rücker et al., 2011). This analytical approach resulted in so-called 'shrunken' discrimination ratios that largely account for small-study publication bias (Harrer et al., 2021; Schwarzer et al., 2020).[13] The differences between the initial discrimination ratios and the robust versions of the discrimination ratios were assessed using *z*-tests (for details on the computational approach, see Altman & Bland, 2003).[14]

Finally, we examined the statistical independence of the sampled discrimination ratios. Interdependency between the discrimination ratios could arise in cases wherein different ratios relied on observations from the same control group (Higgins et al., 2019). For example, if a given experiment consisted of an unmatched design with two distinct treatment groups A and B and one control group C, both $\widehat{DR}_{\text{A–C}}$ (i.e. the discrimination ratio comparing A with C) and $\widehat{DR}_{\text{B–C}}$ (i.e. the discrimination ratio comparing B with C) would be partly based on identical information related to the control group. This factor could lead to the underestimation of between-study variability, which could, in turn, result in false-positive pooled discrimination ratios. To examine this statistical independence, we fitted three-level mixed models including estimates of between-study and within-study heterogeneity with two-level models that only included estimates of within-study heterogeneity per treatment group and compared these models using ANOVA (for the computational approach, see Harrer et al., 2021). We found no evidence that the three-level models had a better fit with the data than the two-level models (see Table A9). We can thus assume that our results were not significantly impacted by interdependency between the sampled discrimination ratios.

[13] Smaller-sized correspondence experiments are, on average, at greater risk of only being reported if they produce large, statistically significant effects vis-à-vis experiments with larger sample sizes (Borenstein et al., 2009; Harrer et al., 2021). Hence, small-study effects can be a source of publication bias due to the correlation between a study's publication status and the nature of its findings.

[14] Here, too, we only calculated the bias statistic and 'shrunken' discrimination ratios when the total number of correspondence tests for a given analysis equalled or exceeded ten. Otherwise, the statistical power would be too low to detect asymmetry, or the statistical heterogeneity of the subset of meta-analyses containing fewer studies would be too high to meaningfully interpret the ratios (see Harrer et al., 2021; Sterne et al., 2011).

3. Results

In this section, we first provide some descriptive statistics regarding the correspondence experiments included in our meta-analysis. Subsequently, we concentrate on the meta-analytic statistics: (i) the pooled discrimination ratios by discrimination ground, (ii) the heterogeneity of these ratios by treatment group, and (iii) their sub-group heterogeneity by call-back classification, region, and period. The statistical heterogeneity (i.e. the statistical measures quantifying between-study variability) and the robustness of the results is discussed alongside the abovementioned statistics. The quasi-exhaustive register of correspondence experiments published between 2005 and 2020, on which our analyses were based, can be retrieved from Table R1.

3.1. Descriptive statistics

Figure 2 shows an increase in the annual number of studies based upon the correspondence testing method published between 2005 and 2020. More specifically, the number of finished experiments rises as of 2005—right after the publication of Bertrand and Mullainathan's (2004) study—and continues to increase steadily in subsequent years.[15] Logically, there is a lag between the year an experiment ends and the year the study is published. On average, this lag is 2.82 years (*SD* = 2.06). While we used the so-called 'Year initially published' for the time-related eligibility criterion in our study selection, the 'Year experiment ended' is used in further analyses because it constitutes a more accurate representation of the timing of a correspondence experiment (see section 3.4.3).

<Figure 2 about here >

In our meta-analysis, we focus on two other grouping variables: region and call-back classification (see section 2.4). Figure 3 represents the number of correspondence experiments (i.e. units of observation) by region. The bulk of correspondence experiments are conducted in Europe (*N* = 196, 64.05%), of which 95 are in Western Europe and 60 in

[15] There is a remarkable peak in the number of publications in 2019. We see two reasons for this sharp increase: (i) many correspondence experiments that ended in previous years (as early as 2013, but mostly in 2016 and 2017) were not published until 2019, and (ii) the *Journal of Ethnic and Migration Studies* compiled a special issue on ethnic discrimination in the labour market that was first published online in 2019.

Northern Europe, and the Americas (*N* = 75, 24.51%), of which 64 are in North America. Figure 4 shows the number of correspondence tests by call-back classification. In the majority of correspondence experiments (*N* = 205, 66.99%), the authors report call-backs in the 'strict' sense (i.e. an invitation to interview), while call-backs in the 'broad' sense (i.e. any positive response from the employer, such as a request for additional information) are reported in 101 experiments (33.01%). A detailed overview of frequencies and proportions by treatment group and region can be found in the appendix (Tables A2 and A3).

< Figure 3 about here >

< Figure 4 about here >

Figure 5 illustrates that the majority of experiments provide results related to the discrimination grounds of race and national origin (*N* = 143, 46.73%) and gender and motherhood status (*N* = 72, 23.53%). Relying on counts, there are two discernible patterns concerning the overall treatment effect. First, for most grounds, there seems to be unequal treatment of applicants from the minority (treatment) group compared with their majority counterparts. Second, the overall treatment of female gender applicants (vis-à-vis male gender applicants) appears highly ambiguous; in the lion's share of the respective experiments (*N* = 33, 53.23%), empirical evidence for unequal treatment is absent. In the following section, we meta-analytically assess these treatment effects per treatment group and address the relevant sub-group differences.

< Figure 5 about here >

3.2. Differences in hiring discrimination by discrimination ground

Unless otherwise indicated, the findings referenced in this section (as well as section 3.3 and 3.4) are (i) robust for controlling for outliers, (ii) robust for funnel plot asymmetry based on the bias statistic, and (iii) equivalent to those obtained in the 'limit' meta-analyses.[16] The detailed results of these robustness analyses can be found in the appendix (Table A4–Table A8). The specific outliers that were removed from the outlier-adjusted statistics can be retrieved from Table A11.

[16] As a reminder, following Harrer et al. (2021) and Sterne et al. (2011), bias statistics were only calculated and 'limit' meta-analyses were only performed for discrimination grounds and treatment groups where $k \geq 10$.

Table 2 includes the pooled discrimination ratios of the correspondence experiments in our meta-analysis. This table also displays these ratios' statistical heterogeneity and heterogeneity by treatment group (see section 3.3). In line with the count of votes in section 3.1, we find convincing empirical evidence for unequal treatment in hiring concerning the discrimination grounds of race and national origin, age, religion, disability, sexual orientation, physical appearance, wealth, and marital status. However, hiring discrimination concerning sexual orientation is not robust when controlling for outliers (*k-adj.* $\widehat{DR}$ = 0.9007, $CI_{95\%}$ = [0.7845; 1.0341]; see Table A4).[17] Moreover, we find no overall evidence of hiring discrimination on the basis of gender and motherhood status or military service or affiliation.

< Table 2 about here >

The pooled discrimination ratios enable us to compare the severity of unequal treatment in hiring across different discrimination grounds. Based on these point estimates, people with disabilities are on average approximately 44% less likely to receive a positive response to a job application ($\widehat{DR}$ = 0.5592, $CI_{95\%}$ = [0.3477; 0.8992]), while estimates on the basis of age and physical appearance indicate average reduced probabilities of a positive response of approximately 40% ($\widehat{DR}$ = 0.5991, $CI_{95\%}$ = [0.5205; 0.6896]) and 37% ($\widehat{DR}$ = 0.6308, $CI_{95\%}$ = [0.4738; 0.8397]), respectively. This contrasts with the discrimination ratios for marital status ($\widehat{DR}$ = 0.8846, $CI_{95\%}$ = [0.8109; 0.9650]) and wealth ($\widehat{DR}$ = 0.8806, $CI_{95\%}$ = [0.8081; 0.9596]), which are significantly different from, yet closer to one. Notably, in recent years, many research efforts have focused on examining hiring discrimination on the basis of race and ethnicity (*k* = 143, 46.73% of total units of observation; $\widehat{DR}$ = 0.6600, $CI_{95\%}$ = [0.6259; 0.6960]). Nonetheless, the unequal treatment of disabled, older, and less physically attractive candidates appears at least equally problematic.

In terms of statistical heterogeneity, we witness high variability in the underlying distribution of true discrimination ratios. I^2 estimates range from 82.36% (age) to 98.53% (sexual orientation)—not considering the exceptional cases of wealth, military service or affiliation, and marital status, which are based on a low number of correspondence

[17] The abbreviation '*k*-adj.' is short for '*k*-adjusted', which indicates that a lower number of studies were included in the analysis, adjusting for influential cases (i.e. outliers). An overview of the specific outliers that were removed (by type of analysis) can be retrieved from Table A11.

experiments. This means that the findings of the experiments clustered within the respective discrimination grounds are highly disparate. Furthermore, most of the 95% prediction intervals are very wide. However, this is not surprising: similar estimates are expected when pooling the discrimination ratios in such broad categories. In the following sections, we assess design-related heterogeneity based on treatment group, call-back classification, region, and period, which helps pinpoint whether this large underlying statistical variability can be (partly) explained by discrepancies among study designs.

3.3. Differences in hiring discrimination by treatment group

In this section, we evaluate the design-related heterogeneity of our findings by treatment group. This approach provides a more granular view of the pooled discrimination ratios described above, as pooling said ratios at the level of the discrimination ground substantially masks relevant information about their underlying variability. Estimates by treatment group are given in Table 2. Figure 6 illustrates the relative change in the probability of a positive call-back for the applicants belonging to the respective treatment groups vis-à-vis their counterparts in the control group (based on the pooled discrimination ratios displayed in Table 2).

< Figure 6 about here >

We find significant differences between the treatment groups clustered within the discrimination grounds of race and national origin, gender and motherhood status, age, religion, and sexual orientation. First, regarding race and national origin, unfavourable treatment in hiring is highest for applicants belonging to the groups Arab/Maghrebi/Middle Eastern ($\widehat{DR}$ = 0.5397, $CI_{95\%}$ = [0.4820; 0.6044]), East Asian/South-East Asian ($\widehat{DR}$ = 0.5681, $CI_{95\%}$ = [0.4225; 0.7639]), and African/African American/Black ($\widehat{DR}$ = 0.6401, $CI_{95\%}$ = [0.5692; 0.7199]; Q = 41.03, p < 0.001), who on average face approximately 46%, 43%, and 36% reductions in the probability of a positive call-back, respectively. We also find overall hiring discrimination against applicants of South Asian or Indian origin ($\widehat{DR}$ = 0.7004, $CI_{95\%}$ = [0.6352; 0.7723]. Furthermore, of all European treatment groups, Southern Europeans experience hiring discrimination to the largest extent ($\widehat{DR}$ = 0.7027, $CI_{95\%}$ = [0.6168; 0.8005]), while the discrimination ratios related to the applicants of (white) Northern and

Western origin ($\widehat{DR}$ = 0.8154, $CI_{95\%}$ = [0.6661; 0.9981]) or Eastern European origin ($\widehat{DR}$ = 0.7206, $CI_{95\%}$ = [0.5271; 0.9851]) are closer to one.[18, 19]

Perhaps more surprisingly, at first, we do not find evidence for overall hiring discrimination against applicants belonging to the Hispanic/Latin American/Caribbean treatment group ($\widehat{DR}$ = 0.8710, $CI_{95\%}$ = [0.7205; 1.0530]), despite many individual correspondence studies that provide evidence for the unequal treatment of this group. However, when the only identified outlier from the analysis of the Hispanic/Latin American/Caribbean treatment group is excluded, the pooled discrimination ratio becomes statistically significant (*k-adj.* $\widehat{DR}$ = 0.8175, $CI_{95\%}$ = [0.7095; 0.9420]; see Table A5). This result is in line with the previous work of Quillian et al. (2017, 2019), who found that discrimination against applicants of Latin American origin seemed to be generally lower than that against applicants belonging to black, Middle Eastern, North African or Asian minority groups.

Next, we take a closer look at the remaining discrimination grounds. We observe no evidence for hiring discrimination related to the treatment groups female gender, mother, or transgender. However, after excluding outliers from the analysis concerning gender, the discrimination ratio for female gender (vis-à-vis male gender) applicants becomes statistically significant and positive (*k-adj.* $\widehat{DR}$ = 1.0663, $CI_{95\%}$ = [1.0221; 1.1124]; see Table A5). Controlling for outliers, we thus find evidence for slightly favourable treatment of female gender candidates in hiring, equal to an overall 6.60% higher probability of receiving a positive response to an application. This finding is confirmed by the estimate of the 'limit' meta-analysis (*limit-adj.* $\widehat{DR}$ = 1.0413, $CI_{95\%}$ = [1.0138; 1.0696]; see Table A6). Upon closer inspection, we notice that this small effect is presumably driven by correspondence experiments performed between 2002 and 2010 and that this effect has disappeared in more recent years (see section 3.4.3). Importantly, we note that (i) the statistical heterogeneity related to this pooled discrimination ratio is high (I^2 = 0.9481, $CI_{95\%}$ = [0.9395;

[18] More specifically, applicants with Albanian-sounding names are discriminated against in Greece and Italy, applicants with Greek names are unfavourably treated in Canada, applicants of Italian origin are discriminated against in Australia and Belgium, and applicants with a Serbian name and appearance experience hiring discrimination in Austria. The control group always consisted of same-country applicants belonging to their region's majority ethnic group.

[19] The Northern and Western European treatment groups comprised minority (majority) applicants of English (Finnish), French (German), German (Irish, Italian, or Russian), and Latvian or Lithuanian (Russian) origin for whom unequal treatment in hiring was assessed. The Eastern European treatment group consisted of minority (majority) applicants of Russian (Finnish), Romanian (Italian), Ukrainian (Russian or Greek), and Polish (Swedish) origin.

0.9556])—exemplified visually in Figure 6—and that (ii) future correspondence experiments are very likely to find effects pointing in opposite directions ($\widehat{DR}^{*}$ $CI_{95\%}$ = [0.6505; 1.6889]). These findings are in line with a recent, large-scale correspondence experiment (including over 80,000 applications) conducted in the United States in which the authors found that contact rates for male gender and female gender applicants differed significantly between companies: some firms favoured male gender candidates, while others favoured female gender candidates (Kline et al., 2021).[20]

Furthermore, older applicants (vis-à-vis younger applicants), but not younger applicants (vis-à-vis older applicants), are strongly discriminated against in correspondence experiments ($\widehat{DR}$ = 0.5804, $CI_{95\%}$ = [0.4993; 0.6748]; Q = 5.64, p = 0.018).[21] However, the estimate for younger applicants is based on a very small number of tests (k = 2). Similarly, hiring discrimination based on religion is mainly driven by the unequal treatment of Muslims ($\widehat{DR}$ = 0.6349, $CI_{95\%}$ = [0.5181; 0.7781]; Q = 16.43, p < 0.001).[22] Discrimination based on disability seems to be largely prompted by the unequal treatment of applicants with a mental disability ($\widehat{DR}$ = 0.6249, $CI_{95\%}$ = [0.4075; 0.9581]), not a physical disability ($\widehat{DR}$ = 0.5369, $CI_{95\%}$ = [0.2607; 1.1056]), although no significant differences between the treatment groups are found (Q = 0.20, p = 0.656). After excluding one outlier from the analysis, however, hiring discrimination based on physical disability also becomes statistically significant (*k-adj.* $\widehat{DR}$ = 0.7494, $CI_{95\%}$ = [0.6259; 0.8974]), while the sub-group difference between the two treatment groups remains insignificant (Q = 1.39, p = 0.239; see Table A5).

[20] It is, however, unclear what exactly drives this variability. For example, demand-side factors, such as the influence of certain job characteristics on the selection criteria used by employers, may lead to gender-based hiring discrimination against either female or male candidates. In addition, the self-selection of members of one gender group into specific sectors, creating a predominance of that gender in those sectors, could lie at the root of additional discrimination against members of the other group.

[21] The operationalisation of age differed substantially across correspondence experiments. We refer the reader to the register of correspondence experiments (Table R1) for details concerning what constitutes 'older' and 'younger' applicants in the original studies. A representative example is the study by Riach (2015), where older candidates were 47 years old, while younger candidates were 27 years old, creating a 20-year age gap between the treatment and the control group.

[22] Notably, in the majority of cases, Muslims are the subject of the correspondence experiments concerning religion (k = 13, 65.00% of total). Other correspondence experiments have focused on a highly diverse subset of religions (i.e. evangelical, Jehovah's Witness, Pentecostal, Christian [generic], Buddhist, Hindu, Jewish, no religious affiliation, and various religious affiliations).

Finally, the results concerning hiring discrimination based on sexual orientation are mixed. The main effect is primarily driven by correspondence experiments considering individuals who have an affiliation with an LGB+ organisation (e.g. membership in an LGBI+ rights organisation; $\widehat{DR}$ = 0.6482, $CI_{95\%}$ = [0.4539; 0.9257]) in comparison with those who directly disclose a same-sex orientation ($\widehat{DR}$ = 1.0585, $CI_{95\%}$ = [0.547; 2.0485], k = 2; Q = 8.75, p = 0.003). The results from the 'limit' meta-analysis confirm this affiliation effect (*limit-adj.* $\widehat{DR}$ = 0.5545, $CI_{95\%}$ = [0.5343; 0.5755]; see Table A6). This finding raises the question of whether hiring discrimination based on sexual orientation is mainly motivated by a discriminatory stance against activism (i.e. affiliation with an organisation that supports LGB+ rights) rather than discriminatory attitudes regarding same-sex orientation per se (see also Baert, 2014). Nonetheless, we must note that, after excluding two outliers from the analysis, the unequal treatment based on affiliation with an LGB+ organisation is only marginally statistically significant (*k-adj.* $\widehat{DR}$ = 0.7924, $CI_{95\%}$ = [0.6203; 1.0122]; see table A5).

3.4. Differences in hiring discrimination by call-back classification, region, and period

In this section, we report the heterogeneity of our results by call-back classification, region, and period. Table 3 contains the results from the sub-group analyses by discrimination ground. We first discuss the most notable difference at the level of the discrimination ground (i.e. related to sexual orientation). The remaining sub-group differences (e.g. related to gender and motherhood status or age) are discussed at the level of the treatment group in the subsequent paragraphs.

At the level of the discrimination ground, we find evidence for significant sub-group differences in unequal treatment related to gender and motherhood status, age, and sexual orientation.[23] Concerning the latter, we find that hiring discrimination against applicants who are affiliated with an LGB+ organisation or who have a same-sex orientation is higher if the call-back is measured and reported in the strict sense (i.e. an invitation to an interview; $\widehat{DR}$ = 0.6534, $CI_{95\%}$ = [0.4540; 0.9404]) rather than the broad sense (i.e. any

[23] In contrast with the findings of Quillian et al. (2019) and Zschirnt and Ruedin (2016), we do not find robust evidence for higher levels of hiring discrimination concerning race and national origin in Europe vis-à-vis the Americas. This discrepancy is presumably because we only looked at differences at the regional level (and not at differences at the country level).

positive response to an application; $\widehat{DR}$ = 0.9970, $CI_{95\%}$ = [0.3606; 2.7565]; Q = 5.52, p = 0.019). Moreover, unfavourable treatment of these candidates appears to be greater in Asia ($\widehat{DR}$ = 0.2489, $CI_{95\%}$ = [0.2218; 0.2793]) than in the Americas (DR = 0.7650, $CI_{95\%}$ = [0.4181; 1.400]) or Europe (DR = 0.7735, $CI_{95\%}$ = [0.5378; 1.1126]; Q = 89.55, p < 0.001). However, we have reasons to believe that these correlations are spurious. First, the discrimination ratio of the Asian region is based on just one study from Cyprus (Drydakis, 2014), which is part of Western Asia according to the United Nations M49 Standard classification to which we adhered. Second, the sub-group differences related to sexual orientation are not robust when controlling for outliers.

< Table 3 about here>

3.4.1. Call-back classification heterogeneity

Table 4 displays the sub-group heterogeneity of the pooled discrimination ratios based on call-back classification, region, and period at the level of the treatment group. We find very little evidence for differences in hiring discrimination by call-back classification. Only in the correspondence experiments where West Asians are considered the treatment group are levels of hiring discrimination higher when call-backs are reported in the strict sense ($\widehat{DR}$ = 0.6916, $CI_{95\%}$ = [0.6023; 0.7941]) compared with studies reporting only call-backs in the broad sense ($\widehat{DR}$ = 0.8078, $CI_{95\%}$ = [0.7332; 0.8900]; Q = 5.25, p = 0.022). However, this finding is not robust if we control for outliers (Q = 2.11, p = 0.147; see Table A8). In other words, levels of hiring discrimination are not significantly different if the authors measure and record call-backs in the strict sense (i.e. an invitation to interview) compared to the broad sense (i.e. any positive response to the application). Hence, the strictness in measuring or reporting call-backs does not seem to relate to the ratio between the probability of a positive call-back for minority applicants and the probability of a positive response for majority-group candidates (i.e. the discrimination ratio).

< Table 4 about here>

3.4.2. Region heterogeneity

We find several regional differences in hiring discrimination. Applicants of West Asian origin (e.g. Azeri, Armenians, Kurds, Uyghurs) experience significantly more discrimination in Asia ($\widehat{DR}$ = 0.5321, $CI_{95\%}$ = [0.4250; 0.6661]) than in Europe ($\widehat{DR}$ = 0.7772, $CI_{95\%}$ = [0.7138; 0.8463]; Q = 33.57, $p < 0.001$).[24] On closer inspection, we notice that the former region comprises both East and West Asian countries (i.e. China, Georgia, and Turkey). These higher levels of hiring discrimination could be explained by (i) the relatively large local presence of these groups in the labour market compared to Europe and (ii) the negative connotations associated with these groups within these particular regions, which do not necessarily exist in European countries (see, e.g. Asali et al., 2018; Maurer-Fazio, 2013).

We find a similar effect for East and South-East Asians (e.g. Chinese, Malaysians) if we consider the outlier-adjusted statistics (the result is only marginally statistically considering the non-adjusted ratios; p = 0.052). These groups also experience more discrimination in Asia (*k-adj.* $\widehat{DR}$ = 0.5941, $CI_{95\%}$ = [0.5269; 0.6699]) than in the Americas (*k-adj.* $\widehat{DR}$ = 0.7842, $CI_{95\%}$ = [0.5385; 1.1369]) or Europe (*k-adj.* $\widehat{DR}$ = 0.6759, $CI_{95\%}$ = [0.5079; 0.8996]; Q = 20.34, p = <0.001; see Table A8). Furthermore, Southern Europeans are treated more unequally in Europe ($\widehat{DR}$ = 0.6553, $CI_{95\%}$ = [0.5653; 0.7621]) compared to the Americas ($\widehat{DR}$ = 0.7615, $CI_{95\%}$ = [0.3710; 1.5629]) and Australia (i.e. the 'other' region; $\widehat{DR}$ = 0.9134, $CI_{95\%}$ = [0.7978; 1.0458]; Q = 12.87, p = 0.002; see Table A8). These last two findings could be explained in an identical manner as the differences for West Asian minorities.

Notably, the regional differences for African, African American, or black applicants also become statistically significant in our outlier-adjusted analysis. Perhaps surprisingly, we find that these groups are more discriminated against in Europe (*k-adj.* $\widehat{DR}$ = 0.5606, $CI_{95\%}$ = [0.4921; 0.6387]) than the Americas (*k-adj.* $\widehat{DR}$ = 0.6771, $CI_{95\%}$ = [0.6128; 0.7481]; Q = 9.53, p = 0.009; see Table A8). Nevertheless, this finding is consistent with the results of the recent meta-study by Thijssen et al. (2021b). They put forward the explanation that because African Americans are perceived as 'culturally native', they enjoy a higher status in American compared to European society and are therefore less discriminated against.

[24] We did not identify any correspondence tests originating from the Americas in which applicants of West Asian origin were considered as the treatment group.

We also observe regional differences in unequal treatment based on age: older applicants are more severely discriminated against in European countries (mainly Belgium, France, and the United Kingdom; $\widehat{DR}$ = 0.5152, $CI_{95\%}$ = [0.4258; 0.6234]) compared with the Americas (i.e. the United States; $\widehat{DR}$ = 0.6916, $CI_{95\%}$ = [0.6342; 0.7541]; Q = 10.23, p = 0.001). This finding is even more exceptional given that the ages in the treatment groups of the European correspondence experiments range from 37 to 56 years, while the ages used in the American studies are generally higher, ranging from 50 to 66 years.[25] Nonetheless, this regional difference is in line with the average employment rate of 55- to 64-year-olds for the period of 2002 (the year of the first correspondence experiments regarding age included in this review) to 2017 (the year of the last correspondence experiments regarding age included in this review) in the United States (60.97%) compared with Belgium (36.89%), France (42.32%), and the United Kingdom (59.89%; OECD, 2021).

3.4.3. Period heterogeneity

We note only one sub-group discrepancy based on the analysis by time period. Female applicants seem to be treated more favourably in correspondence experiments that ended between 2002 and 2010 ($\widehat{DR}$ = 1.1442, $CI_{95\%}$ = [1.0577; 1.2377]) compared with the period from 2011 to 2020 ($\widehat{DR}$ = 0.9984, $CI_{95\%}$ = [0.9130; 1.0919]; Q = 5.50, p = 0.019). Nevertheless, this difference becomes statistically insignificant when the identified outliers are excluded from the analysis (Q = 2.69, p = 0.101).

However, if time is treated as a continuous variable, additional correlation analyses reveal that levels of hiring discrimination related to race and national origin do seem to be in decline. The correlation between (i) the weighted response ratios of the individual correspondence experiments related to race and national origin and (ii) the years in which the respective experiments ended is negative and small but statistically significant ($\hat{r}_{Pearson}$ = −0.25, N_{pairs} = 143; β = −0.0117, t = −3.05, p = 0.003; see Table A10). This equates to an average decrease in hiring discrimination regarding race and ethnicity of 16.42 pp. between 2006 ($\widehat{DR}$ = 0.6054) and 2020 ($\widehat{DR}$ = 0.7696).

[25] As a reminder, Table R1 includes details about the age of the candidates in the treatment and the control groups in the related correspondence experiments.

At first glance, this finding seems to contradict the meta-studies of Heath and Di Stasio (2019) and Quillian et al. (2017), who found few to no temporal differences in hiring discrimination based on race in the United States and the United Kingdom labour markets. However, we note that this correlation is mainly driven by the moderate correlation related to European correspondence experiments ($\hat{r}_{Pearson}$ = −0.40, N_{pairs} = 94; β = −0.0164, t = −4.13, p = <0.001)—averaging a decrease of 23.01 pp. in ethnicity-based hiring discrimination between 2006 ($\widehat{DR}$ = 0.5588) and 2020 ($\widehat{DR}$ = 0.7889)—as opposed to studies conducted in the Americas, where no significant temporal change is observed ($\hat{r}_{Pearson}$ = 0.05, N_{pairs} = 38; β = 0.0033, t = 0.31, p = 0.755; see Table A10 and Figure A1, A2, and A3). Finally, based on the correlation analyses, we find no statistically significant evidence for a change in hiring discrimination in recent years regarding the remaining discrimination grounds.[26]

4. Conclusion

In this meta-analysis, we extensively documented and synthesised the recent hiring discrimination literature grounded in the correspondence testing method, which benefits both scholars and policymakers. Based on research from around the world, we quantified the level of hiring discrimination for ten grounds based on which unequal treatment is forbidden under United States federal or state law: (i) race and national origin, (ii) gender and motherhood status, (iii) religion, (iv) disability, (v) age, (vi) military service or affiliation, (vii) wealth, (viii) marital status, (ix) sexual orientation, and (x) physical appearance. Moreover, we assessed sub-group differences according to how the call-back variable was measured and reported, the region linked to the correspondence experiment, and the related period. Our study provides scholars and policymakers with a broad basis for comparison regarding international research on hiring discrimination. Knowing which and to what extent minority groups face labour market inaccessibility is invaluable in tackling

[26] It must be noted that the correlation coefficients regarding religion, related to European correspondence experiments ($\hat{r}_{Pearson}$ = −0.50, N_{pairs} = 14; p = 0.067), and sexual orientation ($\hat{r}_{Pearson}$ = −0.54, N_{pairs} = 12; p = 0.071) are high but statistically insignificant. These figures might reflect real temporal declines in hiring discrimination, yet we cannot rule out the null hypothesis (at the 5% significance level) that this is not the case.

this issue. Our study also offers insights into relevant contextual variations in hiring discrimination. In the following paragraphs, we first discuss the limitations of our research, followed by the most notable results of our meta-analysis and implications for practice and future work.

Notwithstanding the important contributions of our review, there are a few limitations with regard to the research methods we applied. First, our research is based on a synthesis of only correspondence experiments, whereas other meta-studies have, for example, also included in-person audits to paint a broader picture of hiring discrimination (e.g. Quillian et al., 2017). However, as we argued in the introduction section, in-person audits face a critical limitation. Specifically, behavioural differences between applicants could have an undesirable influence on an employer's assessment in a selection context and therefore muddle the relationship between the individual characteristics of interest (e.g. national origin) and the hiring decision. Second, our meta-analysis might suffer from publication bias to some extent because we did not consider unpublished manuscripts or non-English research. Nonetheless, we statistically evaluated and attempted to control for said bias. Third, we did not perform meta-regressions, in part because the scope of our review was already very broad but also because many relevant covariates were missing at the study level. We advise future studies to apply meta-regressions to sufficiently specific research problems regarding a limited selection of minority groups. This approach would enable scholars to more precisely attribute the uncovered variance of the pooled discrimination ratios to relevant factors beyond the sub-group differences discussed in this review.

We observed five notable results with significant policy and future research implications. Our first observation relates to hiring discrimination at the level of the discrimination ground. Historically, research efforts have focused heavily on examining hiring discrimination based on race and national origin. This research commitment is not unjustified: applicants with salient racial or ethnic characteristics that are significantly different from those of the majority group(s) in a given country are substantially less likely to receive positive responses to their applications. However, it appears that the unequal treatment of applicants with disabilities, older applicants, and less physically attractive applicants is equally problematic. In addition, we found convincing (albeit more modest) evidence of hiring discrimination based on religion, wealth, and marital status. Public and private diversity policies, such as outreach campaigns and diversity training, and other remedial measures should also focus on candidates from these minority groups.

Second, levels of hiring discrimination against the individual treatment groups within the set of examined discrimination grounds generally differ substantially. For example, candidates of Arab, Maghrebi, or Middle Eastern origin are severely discriminated against in the hiring process, facing an estimated average reduced chance of a positive response of about 46%. At the same time, there is only weak evidence of discrimination against (white) European minority applicants. Therefore, measures to decrease hiring discrimination should be targeted at specific relevant minority groups. In this respect, our meta-analysis offers an account of the severity of hiring discrimination against a multitude of minority groups.

Third, hiring discrimination related to sexual orientation is mainly driven by the unequal treatment of candidates who signal an LGB+ organisation affiliation, not candidates who disclose a same-sex orientation. However, it remains unclear to what extent the activism component inherently attached to signalling an LGB+ organisation affiliation causes this unequal treatment. Future research could consider alternating between the two signals within the same institutional context, which could subsequently expose the underlying drivers of hiring discrimination linked to sexual orientation.

Fourth, we found more hiring discrimination against older applicants (vis-à-vis younger applicants) in Europe (i.e. Belgium, France, the United Kingdom, Spain, and Sweden; approximately 49% fewer positive call-backs, on average) versus the United States (approximately 31% fewer positive call-backs, on average). This finding is in line with the historic employment rates of 55- to 64-year-olds in the respective countries. Future studies could look into the specific mechanisms that drive these regional differences.

Fifth, we observed few differences in hiring discrimination over time. However, we did notice that hiring discrimination based on race and national origin has significantly decreased in European correspondence experiments: on average, hiring discrimination based on ethnicity declined by approximately 23 pp. between 2006 and 2020. This finding contrasts with our results related to American correspondence experiments and the results from other meta-studies synthesising audit studies from the United States and the United Kingdom, where (almost) no decline in ethnic hiring discrimination was found (i.e. Heath & Di Stasio, 2019; Quillian et al., 2017). Nevertheless, we did not find evidence for any structural temporal changes in unequal treatment in hiring related to the remaining discrimination grounds within the scope of this review. Overall, hiring discrimination remains a pervasive issue.

References

Altman, D. G. (2003). Statistics notes: Interaction revisited: The difference between two estimates. *BMJ, 326*(7382), 219. https://doi.org/10.1136/bmj.326.7382.219

Altonji, J. G., & Blank, R. M. (1999). Race and gender in the labor market. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, pp. 3143–3259). Elsevier. https://doi.org/10.1016/S1573-4463(99)30039-0

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. https://doi.org/10.1016/j.jce.2018.09.001

Baert, S. (2014). Career lesbians. Getting hired for not having kids? *Industrial Relations Journal, 45*(6), 543–561. https://doi.org/10.1111/irj.12078

Baert, S. (2018). Hiring discrimination: An overview of (almost) all correspondence experiments since 2005. In S. M. Gaddis (Ed.), *Audit studies: Behind the scenes with theory, method, and nuance* (pp. 63–77). Springer. https://doi.org/10.1007/978-3-319-71153-9_3

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Balestra, C., & Fleischer, L. (2018). *Diversity statistics in the OECD: How do OECD countries collect data on ethnic, racial and indigenous identity?* (Organisation for Economic Cooperation and Development [OECD] Statistics Working Papers No. 2018/09). Organisation for Economic Cooperation and Development. https://doi.org/10.1787/89bae654-en

Bartkoski, T., Lynch, E., Witt, C., & Rudolph, C. (2018). A meta-analysis of hiring discrimination against Muslims and Arabs. *Personnel Assessment and Decisions*, *4*(2), 1–16. https://doi.org/10.25035/pad.2018.02.001

Beam, E. A., Hyman, J., & Theoharides, C. (2020). The relative returns to education, experience, and attractiveness for young workers. *Economic Development and Cultural Change*, *68*(2), 391–428. https://doi.org/10.1086/701232

Bertrand, M., & Duflo, E. (2016). *Field experiments on discrimination* (National Bureau of Economic Research [NBER] Working Paper No. 22014). National Bureau of Economic Research. https://doi.org/10.3386/w22014

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*(4), 991–1013. https://doi.org/10.1257/0002828042002561

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, *8*(4), 436–455. https://doi.org/10.2307/144855

Booth, A. L., Leigh, A., & Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics*, *74*(4), 547–573. https://doi.org/10.1111/j.1468-0084.2011.00664.x

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386

Borjas, G. (2020). Labor market discrimination. In G. Borjas (Ed.), *Labor economics* (8th ed., pp. 299–340). McGraw-Hill.

Carlsson, M., & Eriksson, S. (2019). Age discrimination in hiring decisions: Evidence from a field experiment in the labor market. *Labour Economics*, *59*, 173–183. https://doi.org/10.1016/j.labeco.2019.03.002

Chowdhury, S., Ooi, E., & Slonim, R. (2020). *Racial discrimination and white first name adoption: Evidence from a correspondence study in the Australian labour market* (IZA Discussion Papers No. 13208). IZA Institute of Labor Economics. https://www.iza.org/publications/dp/13208/racial-discrimination-and-white-first-name-adoption-evidence-from-a-correspondence-study-in-the-australian-labour-market

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, *10*(4), 417–451. https://doi.org/10.2307/3001616

Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2015). Race and gender effects on employer interest in job applicants: New evidence from a resume field experiment. *Applied Economics Letters*, *23*(12), 853–856. https://doi.org/10.1080/13504851.2015.1114571

Di Stasio, V., Lancee, B., Veit, S., & Yemane, R. (2019). 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. https://doi.org/10.1080/1369183x.2019.1622826

Drydakis, N. (2014). Sexual orientation discrimination in the Cypriot labour market: Distastes or uncertainty? *International Journal of Manpower*, *35*(5), 720–744. https://doi.org/10.1108/ijm-02-2012-0026

Drydakis, N. (2017). Measuring labour differences between natives, non-natives, and natives with an ethnic-minority background. *Economics Letters*, *161*, 27–30. https://doi.org/10.1016/j.econlet.2017.08.031

European Commission. (2021). *Guidance note on the collection and use of equality data based on racial or ethnic origin*. Publications Office of the European Union. https://ec.europa.eu/info/sites/default/files/guidance_note_on_the_collection_and_use_of_equality_data_based_on_racial_or_ethnic_origin.pdf

Flage, A. (2019). Discrimination against gays and lesbians in hiring decisions: A meta-analysis. *International Journal of Manpower, 41*(6), 671–691. https://doi.org/10.1108/ijm-08-2018-0239

Gaddis, S. M. (2014). Discrimination in the credential society: An audit study of race and college selectivity in the labor market. *Social Forces*, *93*(4), 1451–1479. https://doi.org/10.1093/sf/sou111

Gaddis, S. M. (2018). An Introduction to audit studies in the social sciences. In S. M. Gaddis (Ed.), *Audit studies: Behind the scenes with theory, method, and nuance* (pp. 3–44). Springer. https://doi.org/10.1007/978-3-319-71153-9_1

Guul, T. S., Villadsen, A. R., & Wulff, J. N. (2019). Does good performance reduce bad behavior? Antecedents of ethnic employment discrimination in public organizations. *Public Administration Review*, *79*(5), 666–674. https://doi.org/10.1111/puar.13094

Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D. D. (2019). *dmetar: Companion R package for the guide 'Doing meta-analysis in R'* (Version 0.0.9000) [Computer software]. http://dmetar.protectlab.org

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347

Havránek, T., Stanley, T. D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W. R., Rost, K., & Aert, R. C. M. (2020). Reporting guidelines for meta-analysis in economics. *Journal of Economic Surveys*, *34*(3), 469–475. https://doi.org/10.1111/joes.12363

Heath, A. F., & Di Stasio, V. (2019). Racial discrimination in Britain, 1969–2017: A meta-analysis of field experiments on racial discrimination in the British labour market. *The British Journal of Sociology*, *70*(5), 1774–1798. https://doi.org/10.1111/1468-4446.12676

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (1st ed.). Wiley. https://doi.org/10.1002/9781119536604

Hoaglin, D. C. (2016). Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. *Statistics in Medicine*, *35*(4), 485–495. https://doi.org/10.1002/sim.6632

IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*(7), e010247. https://doi.org/10.1136/bmjopen-2015-010247

Jacquemet, N., & Yannelis, C. (2012). Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market. *Labour Economics*, *19*(6), 824–832. https://doi.org/10.1016/j.labeco.2012.08.004

Kline, P., Rose, E., & Walters, C. (2021). *Systemic discrimination among large U.S. employers* (SSRN Electronic Journal Working Paper No. 2021–94). University of Chicago, Becker Friedman Institute for Economics. https://doi.org/10.2139/ssrn.3898669

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, *10*(1), 83–98. https://doi.org/10.1002/jrsm.1316

Larsen, E. N., & Di Stasio, V. (2019). Pakistani in the UK and Norway: Different contexts, similar disadvantage. Results from a comparative field experiment on hiring discrimination. *Journal of Ethnic and Migration Studies*, *47*(6), 1201–1221. https://doi.org/10.1080/1369183x.2019.1622777

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology, 51*(3), 485–504. https://doi.org/10.1002/ejsp.2752

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*(4), 719–748. https://doi.org/10.1093/jnci/22.4.719

Maurer-Fazio, M., & Lei, L. (2015). As rare as a panda. *International Journal of Manpower*, *36*(1), 68–85. https://doi.org/10.1108/ijm-12-2014-0258

Morning, A. (2008). Ethnic classification in global perspective: A cross-national survey of the 2000 census round. *Population Research and Policy Review*, *27*(2), 239–272. https://doi.org/10.1007/s11113-007-9062-5

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, *56*(3), 799–866. https://doi.org/10.1257/jel.20161309

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693–709. https://doi.org/10.2307/2525981

Organisation for Economic Cooperation and Development. (2020a). *All hands in? Making diversity work for all*. https://doi.org/10.1787/efb14583-en

Organisation for Economic Cooperation and Development. (2020b). *International migration outlook 2020*. https://doi.org/10.1787/0c0cc42a-en

Organisation for Economic Cooperation and Development. (2021). *Employment rate by age group* [Data set]. https://doi.org/10.1787/084f32c7-en

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Pager, D. (2016). Are firms that discriminate more likely to go out of business? *Sociological Science*, *3*, 849–859. https://doi.org/10.15195/v3.a36

Patacchini, E., Ragusa, G., & Zenou, Y. (2014). Unexplored dimensions of discrimination in Europe: Homosexuality and physical appearance. *Journal of Population Economics*, *28*(4), 1045–1073. https://doi.org/10.1007/s00148-014-0533-9

Paule, R. C., & Mandel, J. (1982). Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards*, *87*(5), 377–385. https://doi.org/10.6028/jres.087.022

Peters, J. L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA*, *295*(6), 676–680. https://doi.org/10.1001/jama.295.6.676

Quillian, L., Heath, A., Pager, D., Midtbøen, A., Fleischmann, F., & Hexel, O. (2019). Do some countries discriminate more than others? Evidence from 97 field experiments of racial discrimination in hiring. *Sociological Science*, *6*, 467–496. https://doi.org/10.15195/v6.a18

Quillian, L., & Midtbøen, A. H. (2021). Comparative perspectives on racial discrimination in hiring: The rise of field experiments. *Annual Review of Sociology*, *47*(1), 391–415. https://doi.org/10.1146/annurev-soc-090420-035144

Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, *114*(41), 10870–10875. https://doi.org/10.1073/pnas.1706255114

Riach, P. A. (2015). A field experiment investigating age discrimination in four European labour markets. *International Review of Applied Economics, 29*(5), 608–619. https://doi.org/10.1080/02692171.2015.1021667

Rich, J. (2014). *What do field experiments of discrimination in markets tell us? A meta-analysis of studies conducted since 2000* (IZA Discussion Papers No. 8584). IZA Institute of Labor Economics. https://www.iza.org/publications/dp/8584/what-do-field-experiments-of-discrimination-in-markets-tell-us-a-meta-analysis-of-studies-conducted-since-2000

Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. A. (Eds.). (1995). The well-built clinical question: A key to evidence-based decisions. *ACP Journal Club, 123*(3), A12. https://doi.org/10.7326/acpjc-1995-123-3-a12

Rücker, G., Schwarzer, G., Carpenter, J. R., Binder, H., & Schumacher, M. (2011). Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics*, *12*(1), 122–142. https://doi.org/10.1093/biostatistics/kxq046

Rücker, G., Schwarzer, G., Carpenter, J. R., & Schumacher, M. (2008). Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Medical Research Methodology*, *8*(1), 79. https://doi.org/10.1186/1471-2288-8-79

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). Heterogeneity and meta-regression. In G. Schwarzer, J. R. Carpenter, & G. Rücker (Eds.), *Meta-analysis with R* (pp. 85–104). Springer. https://doi.org/10.1007/978-3-319-21416-0_4

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2020). *Metasens: Advanced statistical methods to model and adjust for bias in meta-analysis* (Version 0.6–0) [Computer software]. https://CRAN.R-project.org/package=metasens

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rucker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011).

Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*, d4002. https://doi.org/10.1136/bmj.d4002

Stone, A., & Wright, T. (2013). When your face doesn't fit: Employment discrimination against people with facial disfigurements. *Journal of Applied Social Psychology*, *43*(3), 515–526. https://doi.org/10.1111/j.1559-1816.2013.01032.x

Thijssen, L., Coenders, M., & Lancee, B. (2021a). Ethnic discrimination in the Dutch labor market: Differences between ethnic minority groups and the role of personal information about job applicants—evidence from a field experiment. *Journal of International Migration and Integration, 22*(3), 1125–1150. https://doi.org/10.1007/s12134-020-00795-w

Thijssen, L., van Tubergen, F., Coenders, M., Hellpap, R., & Jak, S. (2021b). Discrimination of black and Muslim minority groups in western societies: Evidence from a meta-analysis of field experiments. *International Migration Review*. Advance online publication. https://doi.org/10.1177/01979183211045044

Thomas, K. (2018). The labor market value of taste: An experimental study of class bias in U.S. employment. *Sociological Science*, *5*, 562–595. https://doi.org/10.15195/v5.a24

United Nations. (2021). *United Nations standard country code*s (Series M: Miscellaneous Statistical Papers No. 49). https://unstats.un.org/unsd/methodology/m49/

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P. T., Langan, D., & Salanti, G. (2015). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*(1), 55–79. https://doi.org/10.1002/jrsm.1164

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Yemane, R., & Fernández-Reino, M. (2019). Latinos in the United States and in Spain: The impact of ethnic group stereotypes on labour market outcomes. *Journal of Ethnic and Migration Studies*, *47*(6), 1240–1260. https://doi.org/10.1080/1369183x.2019.1622806

Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests, 1990–2015. *Journal of Ethnic and Migration Studies*, *42*(7), 1115–1134. https://doi.org/10.1080/1369183X.2015.1133279

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data and code availability

The data used in this study are available at the following URL: https://doi.org/10.34740/KAGGLE/DSV/2709395. The code files are available upon reasonable request.

Declaration of competing interest

There are no relevant financial or non-financial competing interests.

Funding

This study was conducted in the context of the EdisTools project. EdisTools is funded by Research Foundation – Flanders (Strategic Basic Research, S004119N).

CRediT authorship contribution statement

LL: Conceptualisation, methodology, formal analysis, investigation, data curation, writing – original draft, visualisation. SV: Investigation, data curation, writing – original draft, writing – review and editing. SB: Conceptualisation, methodology, data curation, writing – review and editing, supervision, funding acquisition.

Tables and figures

Table 1. Eligibility criteria for study inclusion

Criterion	Details
Study type	Correspondence experiment in which applications were sent in response to vacancies.
Population	(Fictitious) applicants from various minority groups and their majority counterparts.
Outcome	Disadvantageous, unequal treatment in the hiring and selection process (i.e. hiring discrimination).
Comparison	Hiring chances of minority applicants compared with those of majority applicants.
Context	Hiring discrimination related to the grounds upon which unequal treatment is forbidden under United States federal or state law (i.e. race and national origin, gender and motherhood status, religion, disability, age, military service or affiliation, wealth, genetic information, citizenship status, marital status, sexual orientation, political orientation, union affiliation, and physical appearance).
Timing	Study first published between 2005 and 2020.

Notes. The framework used to define the eligibility criteria is based on the PICO (Population, Intervention, Comparison, Outcome) framework first coined by Richardson et al. (1995).

Table 2. Main meta-analytic results by discrimination ground and treatment group

Variable		Effect		Statistical heterogeneity		Treatment group heterogeneity
Discrimination ground or treatment group	k	$\widehat{DR}$ [$CI_{95\%}$]	t (p)	I^2 [$CI_{95\%}$]	$\widehat{DR}^*$ $CI_{95\%}$	Q (p)
Race and national origin	***143***	***0.6600 [0.6259; 0.6960]***	***−15.47*** (<0.001)***	***0.9013 [0.8883; 0.9127]***	***[0.3707; 1.1750]***	***41.03*** (<0.001)***
Arab/Maghrebi/Middle Eastern	31	0.5397 [0.4820; 0.6044]	−11.13*** (<0.001)	0.8737 [0.8317; 0.9052]	[0.3019; 0.9648]	
African/African American/Black	26	0.6401 [0.5692; 0.7199]	−7.83*** (<0.001)	0.8844 [0.8429; 0.9149]	[0.3760; 1.0899]	
West Asian	17	0.7224 [0.6451; 0.8088]	−6.10*** (<0.001)	0.6813 [0.4728; 0.8073]	[0.4895; 1.0660]	
East Asian/South-East Asian	11	0.5681 [0.4225; 0.7639]	−4.25** (0.002)	0.9306 [0.8950; 0.9541]	[0.2075; 1.5553]	
Hispanic/Latin American/Caribbean	10	0.8710 [0.7205; 1.0530]	−1.65 (0.134)	0.8052 [0.6507; 0.8914]	[0.4988; 1.5210]	
Southern European	10	0.7027 [0.6168; 0.8005]	−6.12*** (<0.001)	0.7697 [0.5761; 0.8749]	[0.4935; 1.0005]	
Mixed/Multiple	8	0.6757 [0.4287; 1.0651]	−2.04 (0.081)	0.8622 [0.7489; 0.9244]	[0.1795; 2.5434]	N/A
South Asian/Indian	8	0.7004 [0.6352; 0.7723]	−8.61*** (<0.001)	0.5316 [0.0000; 0.7892]	[0.5642; 0.8695]	
Northern European/Western European	8	0.8154 [0.6661; 0.9981]	−2.39* (0.048)	0.7649 [0.5311; 0.8821]	[0.4738; 1.4031]	
Asian (generic)	5	0.6739 [0.4530; 1.0024]	−2.76 (0.051)	0.7777 [0.4645; 0.9077]	[0.2776; 1.6358]	
Eastern European	5	0.7206 [0.5271; 0.9851]	−2.91* (0.044)	0.9231 [0.8504; 0.9605]	[0.3122; 1.6635]	
Indigenous	3	0.7793 [0.4127; 1.4715]	−1.69 (0.233)	0.9571 [0.9065; 0.9803]	[0.0199; 30.5504]	
Central Asian	1	N/A	N/A	N/A	N/A	
Gender and motherhood status	***72***	***1.0263 [0.9665; 1.0897]***	***0.86 (0.392)***	***0.9407 [0.9311; 0.9489]***	***[0.6454; 1.6319]***	***18.71*** (<0.001)***
Female gender	62	1.0482 [0.9817; 1.1191]	1.44 (0.156)	0.9481 [0.9395; 0.9556]	[0.6505; 1.6889]	
Mother	8	0.9044 [0.7887; 1.0370]	−1.74 (0.126)	0.3049 [0.0000; 0.6902]	[0.6806; 1.2017]	N/A
Transgender	2	0.8500 [0.5306; 1.3619]	−4.38 (0.143)	N/A	N/A	
Age	***19***	***0.5991 [0.5205; 0.6896]***	***−7.65*** (<0.001)***	***0.8236 [0.7352; 0.8825]***	***[0.3465; 1.0357]***	***5.64* (0.018)***
Older age	17	0.5804 [0.4993; 0.6748]	−7.66*** (<0.001)	0.8353 [0.7487; 0.8920]	[0.3331; 1.0114]	N/A
Younger age	2	0.7698 [0.2294; 2.5830]	−2.75 (0.222)	N/A	N/A	
Religion	***21***	***0.6919 [0.5899; 0.8115]***	***−4.82*** (<0.001)***	***0.9245 [0.898; 0.9442]***	***[0.3463; 1.3824]***	***16.43*** (<0.001)***
Muslim	14	0.6349 [0.5181; 0.7781]	−4.82*** (<0.001)	0.8573 [0.7765; 0.9089]	[0.3082; 1.3083]	
Other	3	0.8240 [0.3578; 1.8979]	−1.00 (0.423)	0.9132 [0.7762; 0.9663]	[0.0066; 103.6074]	N/A
Christian	2	0.7293 [0.0075; 71.1483]	−0.88 (0.542)	N/A	N/A	
Multiple	2	0.9275 [0.7532; 1.1422]	−4.59 (0.137)	N/A	N/A	
Disability	***13***	***0.5592 [0.3477; 0.8992]***	***−2.67* (0.021)***	***0.9680 [0.9569; 0.9763]***	***[0.0967; 3.2345]***	***0.20 (0.656)***
Physical disability	9	0.5369 [0.2607; 1.1056]	−1.99 (0.082)	0.9783 [0.9701; 0.9843]	[0.0530; 5.4416]	N/A
Mental disability	4	0.6249 [0.4075; 0.9581]	−3.50* (0.039)	0.3843 [0.0000; 0.7895]	[0.2498; 1.5630]	
Sexual orientation	***12***	***0.7016 [0.5138; 0.9581]***	***−2.50* (0.029)***	***0.9853 [0.9813; 0.9885]***	***[0.2292; 2.1481]***	***8.74** (0.003)***

LGB+ organisation affiliation	10	0.6482 [0.4539; 0.9257]	−2.75* (0.022)	0.9878 [0.9844; 0.9905]	[0.1973; 2.1301]	N/A
Same-sex orientation	2	1.0585 [0.5470; 2.0485]	1.09 (0.471)	N/A	N/A	
Physical appearance	*9*	*0.6308 [0.4738; 0.8397]*	*−3.71** (0.006)*	*0.9788 [0.9708; 0.9846]*	*[0.2615; 1.5214]*	*N/A*
Wealth	*7*	*0.8806 [0.8081; 0.9596]*	*−3.62* (0.011)*	*0.0000 [0.0000; 0.7081]*	*[0.8046; 0.9638]*	*N/A*
Military service or affiliation	*4*	*0.9983 [0.7766; 1.2834]*	*−0.02 (0.985)*	*0.6743[0.0509; 0.8882]*	*[0.5513; 1.8080]*	*N/A*
Marital status	*4*	*0.8846 [0.8109; 0.9650]*	*−4.49* (0.021)*	*0.0000 [0.0000; 0.8469]*	*[0.7865; 0.9950]*	*N/A*

Notes. Abbreviations and notations used: *k* (number of correspondence experiments), $\widehat{DR}$ (pooled discrimination ratio estimate), $CI_{95\%}$ (95% confidence interval), $\widehat{DR}^*$ $CI_{95\%}$ (95% prediction interval of the pooled discrimination ratio), LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations), and N/A (not applicable). Following Schwarzer et al. (2015), treatment group heterogeneity statistics are only calculated for discrimination grounds for which $k \geq 10$. Pooled discrimination rates are only calculated for discrimination grounds or treatment groups for which $k > 1$, while statistical heterogeneity statistics are calculated for those grounds or groups for which $k > 2$. Following Higgins and Thompson (2002), I^2 values around 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3. Sub-group heterogeneity of pooled discrimination ratios by discrimination ground

Variable				Effect		Statistical heterogeneity		Sub-group heterogeneity
Discrimination ground	Sub-group	Level	k	$\widehat{DR}$ [$CI_{95\%}$]	t (p)	I^2 [$CI_{95\%}$]	$\widehat{DR}^*$ $CI_{95\%}$	Q (p)
Race and national origin	Call-back	Stricto	87	0.6592 [0.6188; 0.7021]	−13.13*** (<0.001)	0.8697 [0.8453; 0.8903]	[0.3936; 1.1039]	0.00 (0.983)
		Lato	56	0.6600 [0.5996; 0.7264]	−8.68*** (<0.001)	0.9243 [0.9092; 0.9370]	[0.3325; 1.3100]	
	Region	Americas	38	0.6999 [0.6223; 0.7872]	−6.15*** (<0.001)	0.8860 [0.8534; 0.9113]	[0.3603; 1.3598]	3.32 (0.345)
		Europe	94	0.6607 [0.6234; 0.7001]	−14.18*** (<0.001)	0.8934 [0.8754; 0.9088]	[0.3991; 1.0936]	
		Asia	6	0.4877 [0.2900; 0.8203]	−3.55* (0.016)	0.9303 [0.8754; 0.9610]	[0.1158; 2.0546]	
		Other	5	0.6275 [0.4115; 0.9570]	−3.07* (0.037)	0.9095 [0.8184; 0.9549]	[0.1989; 1.9798]	
	Period	2002–2010	45	0.6476 [0.6033; 0.6952]	−12.35*** (<0.001)	0.7842 [0.7155; 0.8364]	[0.4359; 0.9623]	0.38 (0.535)
		2011–2020	98	0.6680 [0.6228; 0.7163]	−11.45*** (<0.001)	0.9157 [0.9028; 0.9269]	[0.3504; 1.2731]	
Gender and motherhood status	Call-back	Stricto	52	1.0388 [0.9693; 1.1133]	1.10 (0.275)	0.8770 [0.8469; 0.9012]	[0.6640; 1.6252]	0.28 (0.596)
		Lato	20	1.0008 [0.8802; 1.1379]	0.01 (0.990)	0.9687 [0.9603; 0.9752]	[0.5730; 1.7479]	
	Region	Americas	10	0.9658 [0.7412; 1.2584]	−0.30 (0.773)	0.9820 [0.9761; 0.9865]	[0.4248; 2.1958]	5.38 (0.146)
		Europe	48	1.0141 [0.9423; 1.0914]	0.38 (0.703)	0.8992 [0.8749; 0.9187]	[0.6363; 1.6163]	
		Asia	11	1.0471 [0.9463; 1.1586]	1.01 (0.335)	0.8037 [0.6577; 0.8875]	[0.7735; 1.4175]	
		Other	3	1.2854 [0.8355; 1.9775]	2.51 (0.129)	0.4628 [0.0000; 0.8417]	[0.1450; 11.3963]	
	Period	2002–2010	23	1.1236 [1.0327; 1.2224]	2.86** (0.009)	0.8253 [0.7477; 0.8790]	[0.7811; 1.6162]	5.76* (0.016)
		2011–2020	49	0.9825 [0.9097; 1.0612]	−0.46 (0.648)	0.9507 [0.9415; 0.9585]	[0.5996; 1.6101]	
Age	Call-back	Stricto	14	0.5874 [0.4819; 0.7160]	−5.81*** (<0.001)	0.8252 [0.7189; 0.8913]	[0.2999; 1.1503]	0.26 (0.607)
		Lato	5	0.6248 [0.5038; 0.7749]	−6.07** (0.004)	0.7404 [0.3551; 0.8955]	[0.3624; 1.0771]	
	Region	Americas	6	0.6881 [0.6438; 0.7354]	−14.44*** (<0.001)	0.2992 [0.0000; 0.7140]	[0.6195; 0.7642]	7.03** (0.008)
		Europe	13	0.5367 [0.4410; 0.6530]	−6.91*** (<0.001)	0.7887 [0.6447; 0.8743]	[0.2860; 1.0069]	
		Asia	N/A	N/A	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	6	0.5460 [0.4149; 0.7185]	−5.67** (0.002)	0.5205 [0.0000; 0.8088]	[0.2967; 1.0047]	1.00 (0.317)
		2011–2020	13	0.6249 [0.5224; 0.7476]	−5.72*** (<0.001)	0.8645 [0.7853; 0.9145]	[0.3393; 1.1512]	
Religion	Call-back	Stricto	11	0.6795 [0.5242; 0.8807]	−3.32** (0.008)	0.9171 [0.8716; 0.9464]	[0.2912; 1.5857]	0.04 (0.836)
		Lato	10	0.7020 [0.5526; 0.8918]	−3.34** (0.009)	0.9291 [0.8901; 0.9543]	[0.3306; 1.4908]	
	Region	Americas	5	0.7535 [0.5108; 1.1117]	−2.02 (0.113)	0.8197 [0.5839; 0.9219]	[0.2911; 1.9508]	0.39 (0.824)
		Europe	14	0.6832 [0.5588; 0.8353]	−4.09** (0.001)	0.9408 [0.9162; 0.9582]	[0.3258; 1.4329]	
		Asia	2	0.6336 [0.0015; 265.6794]	−0.96 (0.513)	0.9274 [0.7566; 0.9784]	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	

	Period	2002–2010	4	0.5633 [0.2738; 1.1588]	−2.53 (0.085)	0.7796 [0.4043; 0.9184]	[0.0741; 4.2825]	1.08 (0.298)
		2011–2020	17	0.7231 [0.6113; 0.8554]	−4.09*** (<0.001)	0.9016 [0.8584; 0.9317]	[0.3727; 1.4031]	
Disability	Call-back	Stricto	11	0.5331 [0.3007; 0.9451]	−2.45* (0.034)	0.9730 [0.9632; 0.9802]	[0.0742; 3.8310]	1.85 (0.174)
		Lato	2	0.7586 [0.4853; 1.1859]	−7.86 (0.081)	N/A	N/A	
	Region	Americas	5	0.7308 [0.5198; 1.0275]	−2.56 (0.063)	0.7041 [0.2466; 0.8838]	[0.3245; 1.6458]	1.45 (0.228)
		Europe	8	0.4722 [0.2110; 1.0569]	−2.20 (0.064)	0.9802 [0.9724; 0.9859]	[0.0402; 5.5454]	
		Asia	N/A	N/A	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	3	0.3159 [0.0054; 18.5733]	−1.22 (0.348)	0.9940 [0.9910; 0.9961]	[0.0000; >1000.0000]	0.67 (0.414)
		2011–2020	10	0.6863 [0.5701; 0.8261]	−4.59** (0.001)	0.6176 [0.2390; 0.8078]	[0.4217; 1.1167]	
Sexual orientation	Call-back	Stricto	10	0.6534 [0.4540; 0.9404]	−2.64* (0.027)	0.9877 [0.9842; 0.9905]	[0.1945; 2.1954]	5.52* (0.019)
		Lato	2	0.9970 [0.3606; 2.7565]	−0.04 (0.976)	0.5161 [0.0000; 0.8779]	N/A	
	Region	Americas	3	0.7650 [0.4181; 1.4000]	−1.91 (0.197)	0.7581 [0.2036; 0.9265]	[0.0304; 19.2721]	89.55*** (<0.001)
		Europe	8	0.7735 [0.5378; 1.1126]	−1.67 (0.139)	0.9778 [0.9686; 0.9843]	[0.2547; 2.3487]	
		Asia	1	0.2489 [0.2218; 0.2793]	−23.67*** (<0.001)	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	5	0.6159 [0.3468; 1.0936]	−2.34 (0.079)	0.9784 [0.9660; 0.9862]	[0.1251; 3.0309]	0.62 (0.430)
		2011–2020	7	0.7718 [0.4758; 1.2520]	−1.31 (0.238)	0.9873 [0.9827; 0.9907]	[0.1877; 3.1742]	

Notes. Abbreviations and notations used: *k* (number of correspondence experiments), $\widehat{DR}$ (pooled discrimination ratio estimate), $CI_{95\%}$ (95% confidence interval), $\widehat{DR}^{*}\ CI_{95\%}$ (95% prediction interval of the pooled discrimination ratio), and N/A (not applicable). 'Stricto' refers to correspondence experiments in which the call-back variable is related to an invitation to a job interview; 'Lato' refers to experiments in which said variable conveys any positive reaction to an application (e.g. an employer's request for additional information). Following Schwarzer et al. (2015), sub-group heterogeneity statistics are only calculated for discrimination grounds for which $k \geq 10$, while statistical heterogeneity statistics are only calculated for those grounds for which $k > 2$. Following Higgins and Thompson (2002), I^2 values around 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4. Sub-group heterogeneity of pooled discrimination ratios by treatment group

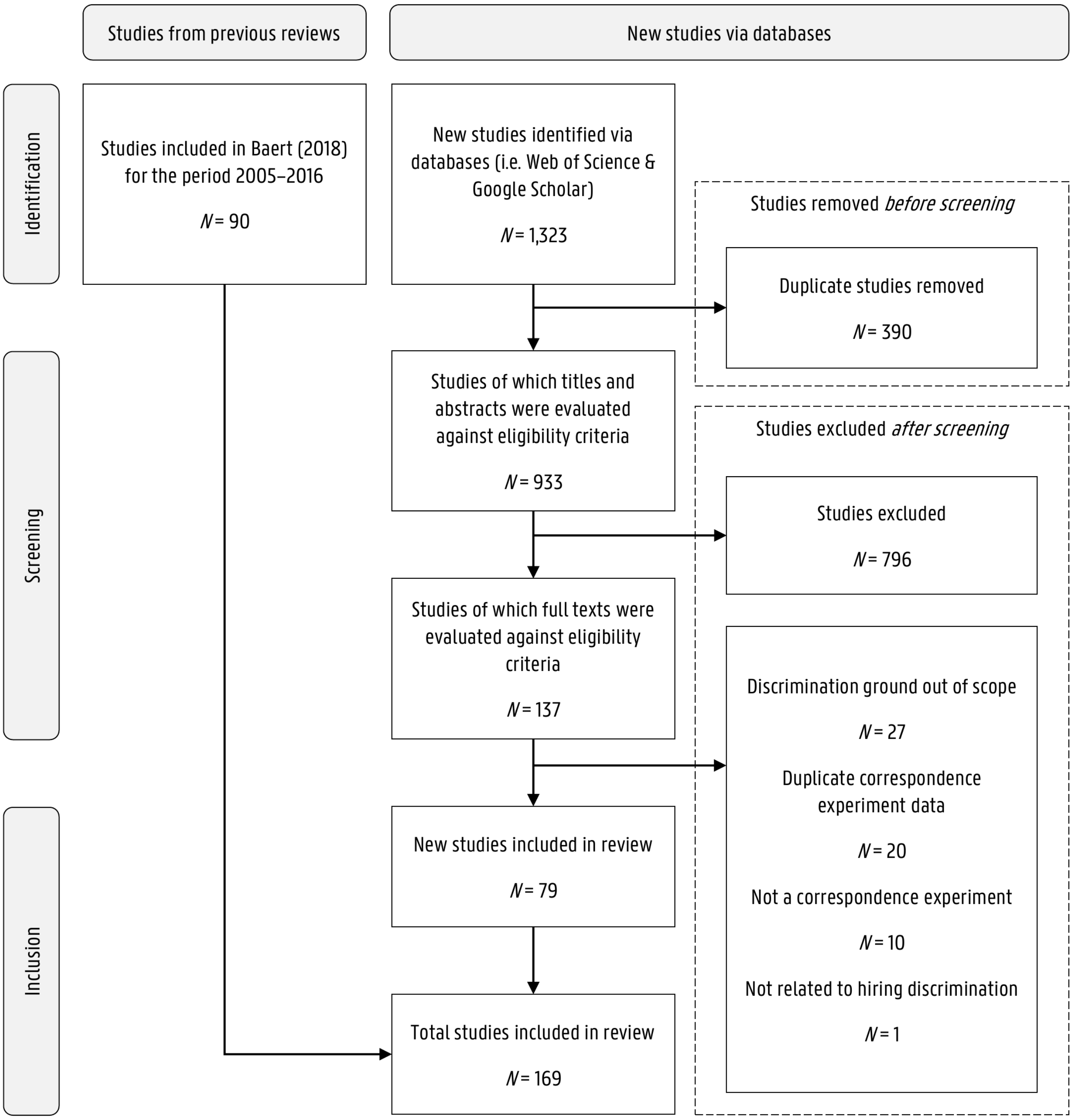
Variable				Effect		Statistical heterogeneity		Sub-group heterogeneity
Discrimination ground or treatment group	Sub-group	Level	*k*	$\widehat{DR}$ [$CI_{95\%}$]	*t* (*p*)	I^2 [$CI_{95\%}$]	$\widehat{DR}^*$ $CI_{95\%}$	*Q* (*p*)
Race and national origin								
Arab/Maghrebi/Middle Eastern	Call-back	Stricto	19	0.5688 [0.4999; 0.6471]	−9.19*** (<0.001)	0.8800 [0.8273; 0.9167]	[0.3385; 0.9558]	1.37 (0.242)
		Lato	12	0.4937 [0.3927; 0.6209]	−6.78*** (<0.001)	0.8743 [0.7989; 0.9214]	[0.2280; 1.0692]	
	Region	Americas	2	0.5967 [0.0430; 8.2789]	−2.49 (0.243)	N/A	N/A	2.74 (0.254)
		Europe	28	0.5339 [0.4721; 0.6038]	−10.47*** (<0.001)	0.8842 [0.8444; 0.9138]	[0.2901; 0.9827]	
		Asia	N/A	N/A	N/A	N/A	N/A	
		Other	1	0.6288 [0.5375; 0.7356]	−5.79*** (<0.001)	N/A	N/A	
	Period	2002–2010	13	0.5587 [0.4697; 0.6646]	−7.31*** (<0.001)	0.8255 [0.7142; 0.8935]	[0.3093; 1.0090]	0.26 (0.612)
		2011–2020	18	0.5282 [0.4492; 0.6211]	−8.31*** (<0.001)	0.8963 [0.8515; 0.9276]	[0.2753; 1.0135]	
African/African American/Black	Call-back	Stricto	15	0.6724 [0.5916; 0.7642]	−6.65*** (<0.001)	0.6332 [0.3609; 0.7895]	[0.4604; 0.9819]	0.49 (0.483)
		Lato	11	0.6185 [0.4914; 0.7784]	−4.65*** (<0.001)	0.9431 [0.9160; 0.9615]	[0.2857; 1.3389]	
	Region	Americas	17	0.6444 [0.5460; 0.7606]	−5.62*** (<0.001)	0.8740 [0.8137; 0.9148]	[0.3429; 1.2112]	2.17 (0.337)
		Europe	8	0.6156 [0.5058; 0.7492]	−5.84*** (<0.001)	0.9202 [0.8665; 0.9523]	[0.3670; 1.0325]	
		Asia	1	0.8000 [0.5874; 1.0895]	−1.42 (0.157)	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	7	0.6162 [0.5268; 0.7208]	−7.56*** (<0.001)	0.3550 [0.0000; 0.7277]	[0.4651; 0.8164]	0.18 (0.671)
		2011–2020	19	0.6423 [0.5506; 0.7492]	−6.04*** (<0.001)	0.9061 [0.8682; 0.9331]	[0.3432; 1.2021]	
West Asian	Call-back	Stricto	14	0.6916 [0.6023; 0.7941]	−5.76*** (<0.001)	0.6623 [0.4065; 0.8078]	[0.4514; 1.0595]	5.25* (0.022)
		Lato	3	0.8078 [0.7332; 0.8900]	−9.48* (0.011)	0.0000 [0.0000; 0.8960]	[0.6069; 1.0753]	
	Region	Americas	N/A	N/A	N/A	N/A	N/A	33.57*** (<0.001)
		Europe	14	0.7772 [0.7138; 0.8463]	−6.40*** (<0.001)	0.3289 [0.0000; 0.6456]	[0.6229; 0.9699]	
		Asia	3	0.5321 [0.4250; 0.6661]	−12.09** (0.007)	0.0000 [0.0000; 0.8960]	[0.2741; 1.0328]	
		Other	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	4	0.6883 [0.4262; 1.1114]	−2.48 (0.089)	0.8552 [0.6433; 0.9412]	[0.1753; 2.7030]	0.14 (0.704)
		2011–2020	13	0.7314 [0.6514; 0.8212]	−5.88*** (<0.001)	0.5927 [0.2495; 0.7790]	[0.5286; 1.0119]	
East Asian/South-East Asian	Call-back	Stricto	7	0.5006 [0.3156; 0.7938]	−3.67* (0.010)	0.9483 [0.9158; 0.9682]	[0.1324; 1.8918]	2.22 (0.136)
		Lato	4	0.6996 [0.4744; 1.0317]	−2.93 (0.061)	0.8624 [0.6648; 0.9435]	[0.2335; 2.0958]	
	Region	Americas	2	0.7824 [0.5385; 1.1369]	−8.34 (0.076)	N/A	N/A	7.73 (0.052)
		Europe	5	0.6759 [0.5079; 0.8996]	−3.80* (0.019)	0.8401 [0.6401; 0.929]	[0.3294; 1.3871]	
		Asia	2	0.3373 [0.0002; 484.3331]	−1.90 (0.308)	N/A	N/A	

		Other	2	0.4689 [0.0182; 12.0734]	−2.96 (0.207)	N/A	N/A	
	Period	2002–2010	4	0.6886 [0.5218; 0.9085]	−4.28* (0.023)	0.5545 [0.0000; 0.8525]	[0.3604; 1.3156]	1.79 (0.181)
		2011–2020	7	0.5160 [0.3183; 0.8364]	−3.35* (0.015)	0.9555 [0.9291; 0.9721]	[0.1256; 2.1199]	
Southern European	Call-back	Stricto	7	0.7234 [0.6156; 0.8501]	−4.91** (0.003)	0.7851 [0.5563; 0.8959]	[0.4898; 1.0685]	0.47 (0.492)
		Lato	3	0.6561 [0.3818; 1.1273]	−3.35 (0.079)	0.7169 [0.0407; 0.9164]	[0.0419; 10.2836]	
	Region	Americas	2	0.7615 [0.3710; 1.5629]	−4.82 (0.130)	N/A	N/A	12.87** (0.002)
		Europe	7	0.6553 [0.5635; 0.7621]	−6.85*** (<0.001)	0.6464 [0.2033; 0.8430]	[0.4618; 0.9299]	
		Asia	N/A	N/A	N/A	N/A	N/A	
		Other	1	0.9134 [0.7978; 1.0458]	−1.31 (0.19)	N/A	N/A	
	Period	2002–2010	5	0.7189 [0.5609; 0.9213]	−3.69* (0.021)	0.8544 [0.6784; 0.9341]	[0.3866; 1.3366]	0.20 (0.655)
		2011–2020	5	0.6814 [0.5455; 0.8510]	−4.79** (0.009)	0.6127 [0.0000; 0.8545]	[0.4282; 1.0842]	
Gender and motherhood status								
Female gender	Call-back	Stricto	44	1.0570 [0.9795; 1.1407]	1.47 (0.149)	0.8919 [0.8640; 0.9141]	[0.6653; 1.6793]	0.11 (0.736)
		Lato	17	1.0292 [0.8882; 1.1927]	0.41 (0.684)	0.9735 [0.9662; 0.9793]	[0.5651; 1.8746]	
	Region	Americas	9	1.0068 [0.7785; 1.3021]	0.06 (0.953)	0.9840 [0.9785; 0.9880]	[0.4622; 2.1930]	4.22 (0.239)
		Europe	38	1.0381 [0.9501; 1.1342]	0.85 (0.398)	0.9168 [0.8954; 0.9338]	[0.6214; 1.7342]	
		Asia	11	1.0471 [0.9463; 1.1586]	1.01 (0.335)	0.8037 [0.6577; 0.8875]	[0.7735; 1.4175]	
		Other	3	1.2854 [0.8355; 1.9775]	2.51 (0.129)	0.4628 [0.0000; 0.8417]	[0.1450; 11.3963]	
	Period	2002–2010	21	1.1442 [1.0577; 1.2377]	3.58** (0.002)	0.8211 [0.7365; 0.8786]	[0.8299; 1.5774]	5.53* (0.019)
		2011–2020	40	0.9962 [0.9090; 1.0918]	−0.08 (0.933)	0.9594 [0.9515; 0.9661]	[0.5786; 1.7151]	
Age								
Older age	Call-back	Stricto	12	0.5573 [0.4464; 0.6956]	−5.80*** (<0.001)	0.8247 [0.7064; 0.8953]	[0.2769; 1.1215]	0.81 (0.368)
		Lato	5	0.6248 [0.5038; 0.7749]	−6.07** (0.004)	0.7404 [0.3551; 0.8955]	[0.3624; 1.0771]	
	Region	Americas	5	0.6916 [0.6342; 0.7541]	−11.83*** (<0.001)	0.4361 [0.0000; 0.7929]	[0.5877; 0.8138]	10.23** (0.001)
		Europe	12	0.5152 [0.4258; 0.6234]	−7.66*** (<0.001)	0.7452 [0.5499; 0.8558]	[0.2911; 0.9117]	
		Asia	N/A	N/A	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	6	0.546 [0.4149; 0.7185]	−5.67** (0.002)	0.5205 [0.0000; 0.8088]	[0.2967; 1.0047]	0.46 (0.497)
		2011–2020	11	0.601 [0.4891; 0.7385]	−5.51*** (<0.001)	0.8811 [0.8069; 0.9268]	[0.3112; 1.1605]	
Religion								
Muslim	Call-back	Stricto	9	0.6683 [0.4933; 0.9054]	−3.06* (0.016)	0.7805 [0.5850; 0.8839]	[0.2709; 1.6491]	0.58 (0.446)
		Lato	5	0.5795 [0.4009; 0.8379]	−4.11* (0.015)	0.9253 [0.8556; 0.9614]	[0.2231; 1.5054]	
	Region	Americas	2	0.7019 [0.0303; 16.2621]	−1.43 (0.388)	N/A	N/A	0.17 (0.918)

		Europe	10	0.6282 [0.4938; 0.7991]	−4.37** (0.002)	0.8689 [0.7785; 0.9224]	[0.2974; 1.3269]	
		Asia	2	0.6336 [0.0015; 265.6794]	−0.96 (0.513)	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	2	0.5840 [0.0003; >1000.0000]	−0.91 (0.531)	N/A	N/A	0.02 (0.882)
		2011–2020	12	0.6385 [0.5250; 0.7765]	−5.05*** (<0.001)	0.8651 [0.7821; 0.9165]	[0.3325; 1.2258]	
Sexual orientation								
LGB+ organisation affiliation	Call-back	Stricto	8	0.5815 [0.3828; 0.8834]	−3.07* (0.018)	0.9903 [0.9874; 0.9926]	[0.1614; 2.0956]	7.71** (0.005)
		Lato	2	0.9970 [0.3606; 2.7565]	−0.04 (0.976)	N/A	N/A	
	Region	Americas	2	0.6654 [0.2679; 1.6523]	−5.69 (0.111)	N/A	N/A	128.56*** (<0.001)
		Europe	7	0.7363 [0.4894; 1.1077]	−1.83 (0.116)	0.9807 [0.9724; 0.9865]	[0.2220; 2.4424]	
		Asia	1	0.2489 [0.2218; 0.2793]	−23.67*** (<0.001)	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	5	0.6159 [0.3468; 1.0936]	−2.34 (0.079)	0.9784 [0.9660; 0.9862]	[0.1251; 3.0309]	0.10 (0.755)
		2011–2020	5	0.6830 [0.3321; 1.4049]	−1.47 (0.216)	0.9914 [0.9880; 0.9939]	[0.0919; 5.0745]	

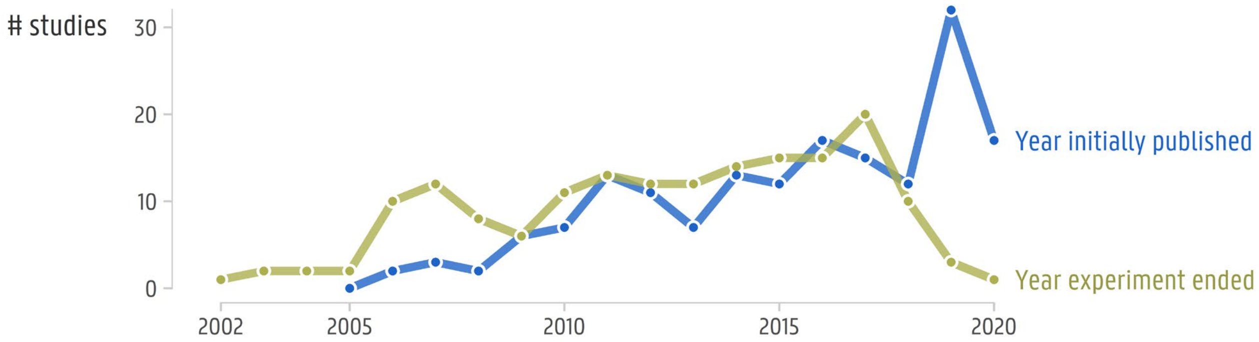
Notes. Abbreviations and notations used: *k* (number of correspondence experiments), $\widehat{DR}$ (pooled discrimination ratio estimate), $CI_{95\%}$ (95% confidence interval), $\widehat{DR}^{*}CI_{95\%}$ (95% prediction interval of the pooled discrimination ratio), N/A (not applicable), and LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations). 'Stricto' refers to correspondence experiments in which the call-back variable is related to an invitation to a job interview; 'Lato' refers to experiments in which said variable conveys any positive reaction to an application (e.g. an employer's request for additional information). Following Schwarzer et al. (2015), sub-group heterogeneity statistics are only calculated for treatment groups for which $k \geq 10$, while statistical heterogeneity statistics are calculated for those groups for which $k > 2$. Following Higgins and Thompson (2002), I^2 values around 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 1. Study selection flow diagram



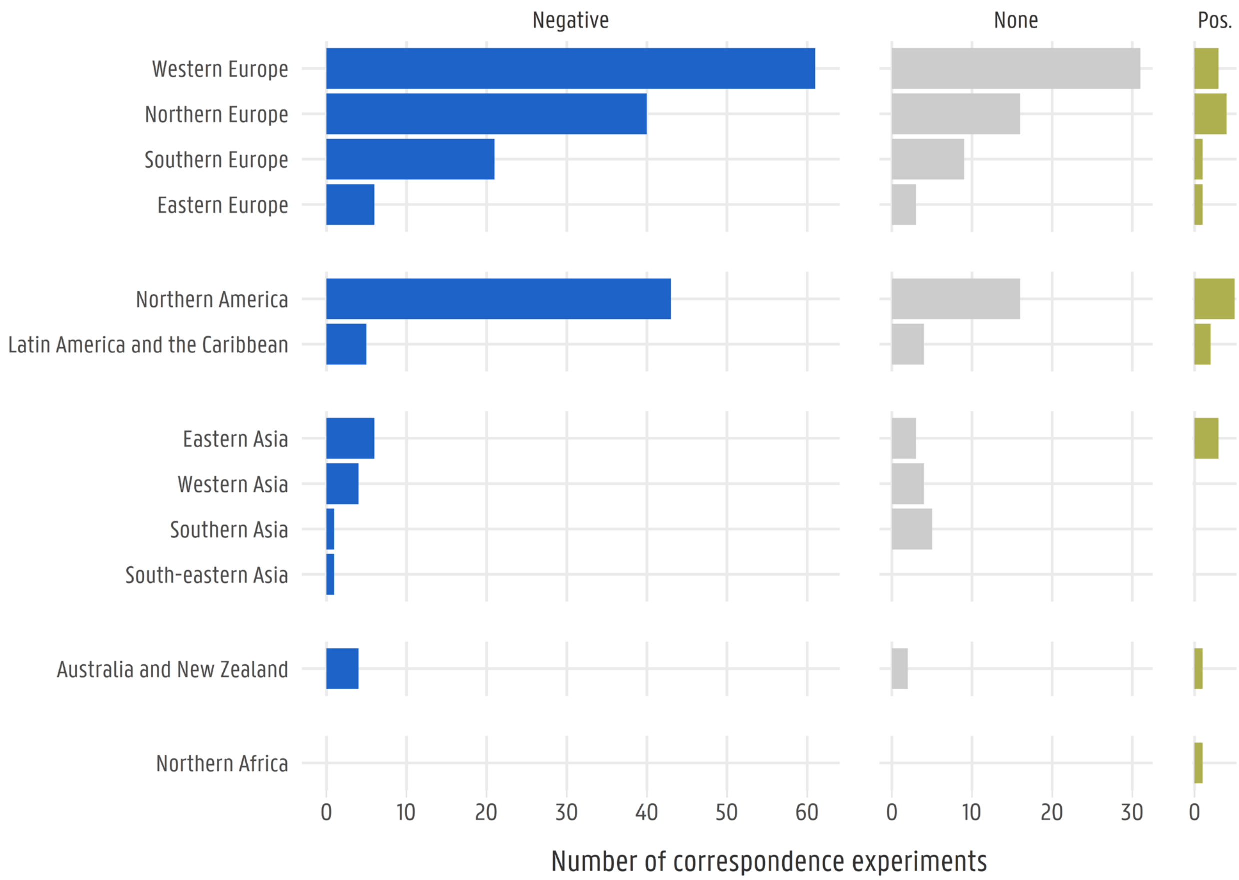
Notes. This figure is adapted from Page et al. (2021, p. 5).

Figure 2. Time trend of the number of studies based upon correspondence experiments



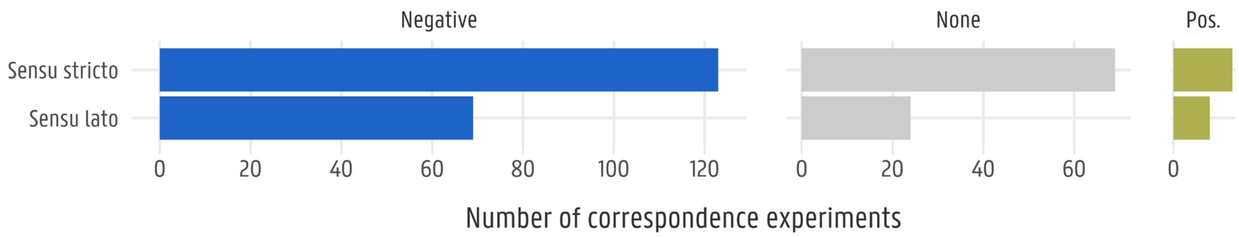
Notes. 'Year initially published' is the year in which the study was first published (as a pre-print, early-access article or a full journal article). This year is used in our research as a criterion for study selection, while 'Year experiment ended' is used in our sub-group analyses as the time period variable.

Figure 3. Number of correspondence experiments by region (rows) and treatment effect (panels)



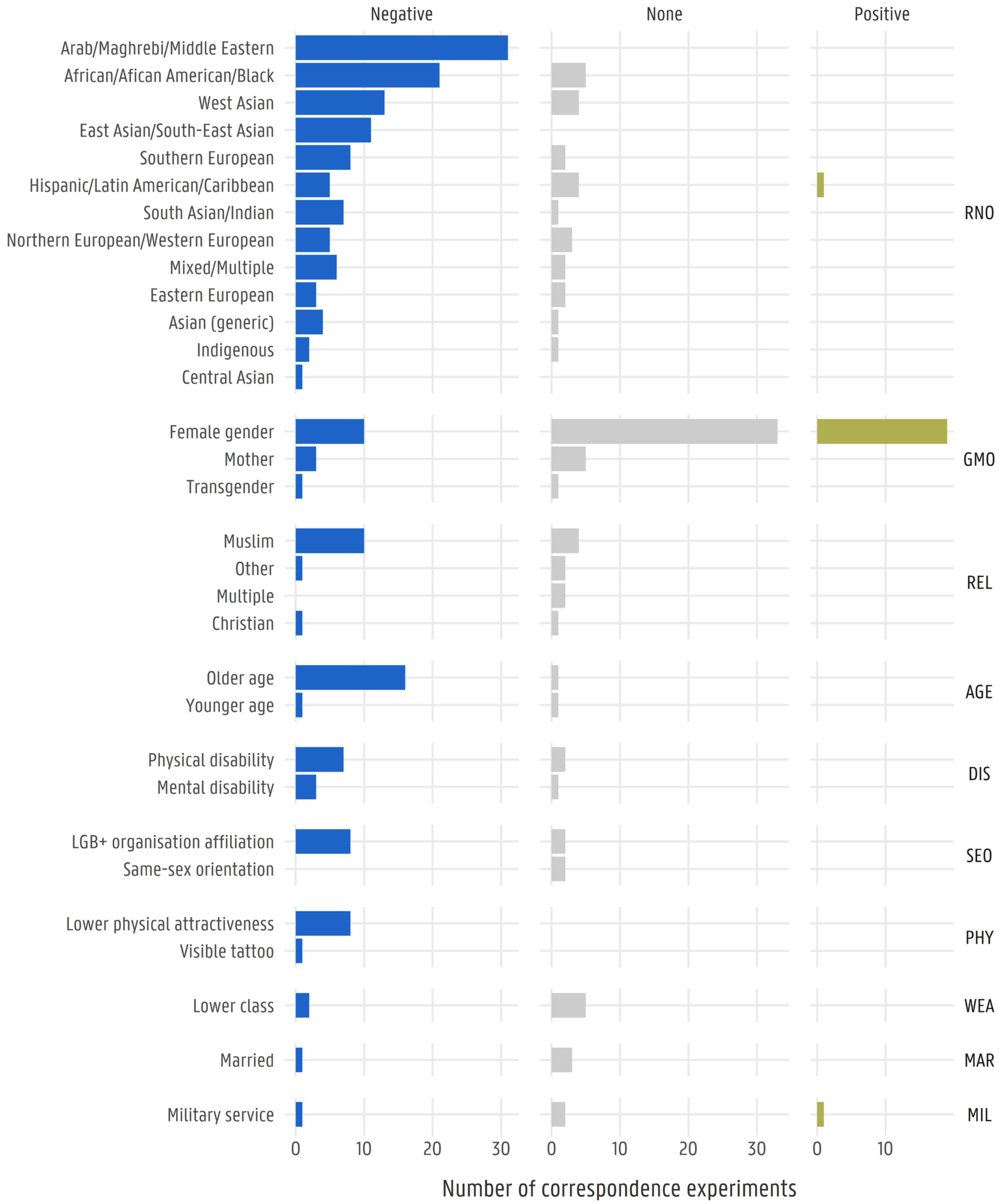
Notes. 'Number of correspondence experiments' represents the units of observation included in the meta-analysis. The bars are grouped in panels, representing the overall treatment effect in the original correspondence experiments. Region classification is based on the United Nations (2021) M49 Standard. Abbreviations used: Pos. (Positive).

Figure 4. Number of correspondence experiments by call-back classification (rows) and treatment effect (panels)



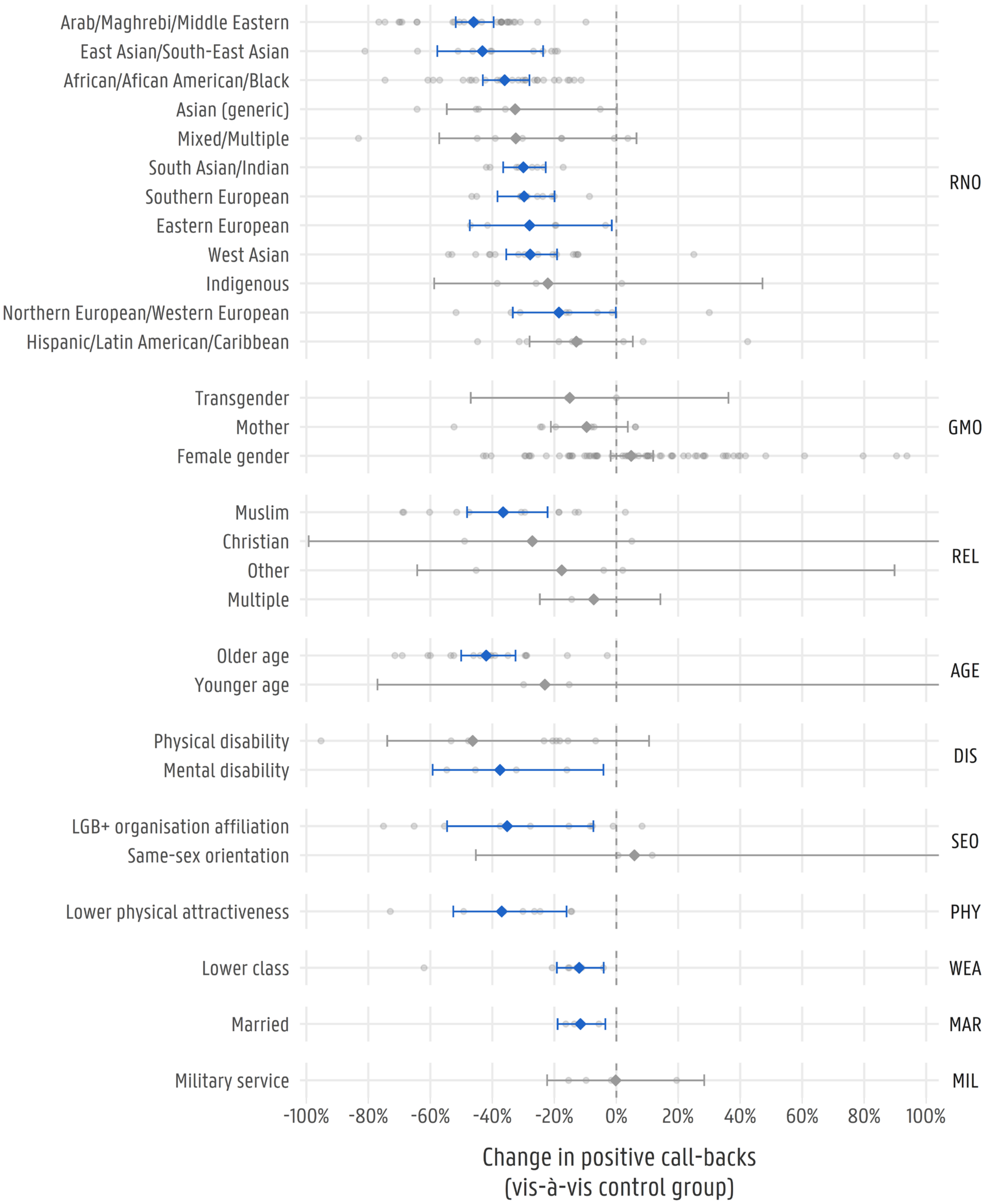
Notes. Outcome variables consisting of an invitation to a job interview (or any broadly defined positive response from the employer, such as a request for additional information) are classified as *sensu stricto* (or *sensu lato*). 'Number of correspondence experiments' represents the units of observation included in the meta-analysis. The bars are grouped in panels, representing the overall treatment effect in the original correspondence experiments. Abbreviations used: Pos. (Positive).

Figure 5. Number of correspondence experiments by treatment group (rows) and treatment effect (panels)



Notes. 'Number of correspondence experiments' represents the units of observation included in the meta-analysis. The bars are grouped in panels, representing the overall treatment effect in the original correspondence experiments. Abbreviations used: RNO (race and national origin), GMO (gender and motherhood status), AGE (age), REL (religion), DIS (disability), SEO (sexual orientation), PHY (physical appearance), WEA (wealth), MIL (military service or affiliation), and MAR (marital status).

Figure 6. Hiring discrimination based on pooled discrimination ratios by treatment group



Notes. The change in positive call-backs (compared to the control group), represented by filled diamond shapes, is calculated by subtracting one from the corresponding pooled discrimination ratio (i.e. $\widehat{DR}$; see Table 2). Error bars illustrate the 95% confidence intervals of these ratios. Statistically insignificant ratios (at the 5% level) are greyed out. Semi-transparent dots represent the discrimination ratios of the individual correspondence experiments. Abbreviations used: RNO (race and national origin), GMO (gender and motherhood status), AGE (age), REL (religion), DIS (disability), SEO (sexual orientation), PHY (physical appearance), WEA (wealth), MIL (military service or affiliation), and MAR (marital status).

Appendix

Table A1. Reporting guidelines for meta-analyses in economics: a checklist

Criterion	Reporting
Research questions and effect size	
A clear statement of the specific effects studied.	In section 1, we describe that our analysis is based on the hiring discrimination ratios produced by correspondence experiments. In section 2, we further delineate the scope of our review and specify the data we collected to quantify these ratios.
A precise definition of how effects are measured and their standard errors, accompanied by any relevant formulas if transformations are made.	In section 2.4, we disclose our empirical strategy, providing formulas for the (pooled) discrimination ratios. We also disclose details on the estimation approach of the standard errors.
An explicit description of how measured effects are comparable, including any methods or formulas used to standardise or convert them to a common metric.	The effects of the individual studies are compared using the corresponding discrimination ratios, which are based on the risk ratio (or relative risk). The method to calculate the standardised pooled discrimination ratios is explicitly described in section 2.3.
Research literature search, compilation, and coding	
A full report of how the research literature was searched.	In section 2.1 and 2.2, we explicitly describe our search and selection strategy.
A full disclosure of the rules for study (or effect size) inclusion/exclusion.	We only use criteria to include studies—these are described in section 2.1. We do not use specific thresholds to include or exclude discrimination ratios.
A statement addressing who searched, read, and coded the research literature.	All studies included in this meta-analysis were independently reviewed by at least two reviewers. Any disagreement was resolved in consensus.
A complete list of the information coded for each study or estimate (i.e. estimated effect size and standard error, at minimum).	An enumeration of the coded variables is provided section 2.3. Moreover, we have included, as an appendix, the complete register of correspondence experiments supplemented with relevant study characteristics.
The rule or method used to identify outliers, leverage, or influence points when omitted.	We disclose the rule for identifying outliers in section 2.4, along with details on the 'limit' meta-analyses we have performed to estimate robust, 'shrunken' discrimination ratio estimates.
Modelling issues	
A table displaying definitions of all the coded variables along with their descriptive statistics (means and standard deviations).	We report the details of the coded variables in section 2.3. In the manuscript, we have included figures depicting count data regarding the main variables. Tabulated overviews of the counts and proportions related to the treatment group classification and the region of the correspondence experiments are provided in Table A2 and A3.
An investigation of publication, selection, and misspecification biases unless these biases can reasonably be expected to be absent.	These biases are examined (and partly controlled for) in our robustness analyses, which include outlier analyses, the computation of a bias statistic to assess funnel plot asymmetry, as well as 'limit' meta-analyses.
Further reporting and interpretation	
Graph(s) of the effect sizes, such as funnel graphs, forest plots, or other statistical displays of data.	We have included a forest plot of the change in the probability of a positive call-back by treatment group (based on the pooled discrimination ratios).
A discussion of the economic (or practical) significance of the main findings.	A discussion (of the importance) of the research aims and the findings starts in section 1 and is concluded in section 4.
A statement about sharing the data or link to its public posting along with the codes of the core analyses.	See the data and code availability statement.

Notes. Criteria from this checklist were adapted from Havránek et al. (2020, p. 471–472).

Table A2. Correspondence experiments included in the meta-analysis by region

Region or sub-region	Country	*N*	Proportion
Americas		*75*	*100.00%*
Northern America	United States of America	55	73.33%
	Canada	9	12.00%
Latin America and the Caribbean	Mexico	5	6.67%
	Peru	3	4.00%
	Argentina	1	1.33%
	Brazil	1	1.33%
	Jamaica	1	1.33%
Asia		*27*	*100.00%*
Eastern Asia	China	12	44.44%
Western Asia	Cyprus	2	7.41%
	Georgia	2	7.41%
	Israel	2	7.41%
	Turkey	2	7.41%
Southern Asia	India	4	14.81%
	Pakistan	2	7.41%
South-Eastern Asia	Malaysia	1	3.70%
Europe		*196*	*100.00%*
Western Europe	Belgium	26	13.27%
	Germany	21	10.71%
	France	19	9.69%
	Netherlands	16	8.16%
	Switzerland	7	3.57%
	Austria	6	3.06
Northern Europe	Sweden	29	14.80%
	United Kingdom of Great Britain and Northern Ireland	16	8.16%
	Norway	5	2.55%
	Finland	4	2.04%
	Denmark	3	1.53%
	Ireland	3	1.53%
Southern Europe	Italy	13	6.63%
	Spain	11	5.61%
	Greece	7	3.57%
Eastern Europe	Russian Federation	8	4.08%
	Czechia	2	1.02%
Other: Africa		*1*	*100.00%*
Northern Africa	Algeria	1	100.00%
Other: Oceania		*7*	*100.00%*
Australia and New Zealand	Australia	7	100.00%

Notes. Counts represent units of observation included in the meta-analysis.

Table A3. Correspondence experiments included in the meta-analysis by treatment group

Treatment group classification (general)	Treatment group classification (specific)	N	Proportion
Race and national origin		*143*	*100.00%*
African/Afican American/Black	Black-sounding name	10	6.99%
	African American name or origin	5	3.50%
	African name	4	2.80%
	Black or dark phenotype	4	2.80%
	Caribbean name	1	0.70%
	Congolese name	1	0.70%
	Nigerian name and appearance	1	0.70%
Arab/Maghrebi/Middle Eastern	Middle Eastern name	10	6.99%
	Moroccan name	8	5.59%
	Arab name or origin	7	4.90%
	North African name	3	2.10%
	Iraqi or Somali name	2	1.40%
	Somali American name	1	0.70%
Asian (generic)	Asian or Asian-sounding name	4	2.80%
	Asian phenotype	1	0.70%
Central Asian	Tajik or Uzbek name	1	0.70%
East Asian/South-East Asian	Chinese name or appearance	7	4.90%
	Malaysian name	1	0.70%
	Mongolian or Tibetan name	1	0.70%
	Southeast or East Asian name	1	0.70%
	Tatar name	1	0.70%
Eastern European	Eastern European or Russian name	2	1.40%
	Polish name	1	0.70%
	Romanian name	1	0.70%
	Ukrainian name	1	0.70%
Hispanic/Latin American/Caribbean	Hispanic or Hispanic-sounding name	7	4.90%
	Cuban, Dominican, Ecuadorian, Mexican, Puerto Rican, or Salvadorian name	2	1.40%
	Antillean name	1	0.70%
Indigenous	Indigenous or indigenous-sounding name or origin	3	2.10%
Mixed/Multiple	Foreign-born	3	2.10%
	Asian domestic-born or white domestic-born	1	0.70%
	Foreign name	1	0.70%
	Mestizo phenotype	1	0.70%
	Roma name	1	0.70%
	Surinamese name	1	0.70%
Northern European/Western European	German name	4	2.80%
	English name	1	0.70%
	French name	1	0.70%
	French or German name	1	0.70%
	Latvian or Lithuanian name	1	0.70%
South Asian/Indian	Indian or Pakistani name	3	2.10%
	Pakistani name	4	2.80%
	South Asian name	1	0.70%
Southern European	Albanian name	3	2.10%
	Greek name	2	1.40%
	Italian name	2	1.40%
	Kosovar name	2	1.40%

	Serbian name and appearance	1	0.70%
West Asian	Turkish name or appearance	13	9.09%
	Armenian, Azeri, Chechen, or Georgian name	1	0.70%
	Azeri or Armenian name	1	0.70%
	Kurdish name	1	0.70%
	Uyghur name	1	0.70%
Gender and motherhood status		*72*	*100.00%*
Female gender	Female gender	62	86.11%
	Female gender with feminine personality traits	1	1.39%
Mother	Mother	7	9.72%
	Pregnant	1	1.39%
Transgender	Transgender	2	2.78%
Religion		*21*	*100.00%*
Muslim	Muslim	13	61.90%
	Wearing a headscarf	1	4.76%
Other	Expressing a religious identity	1	4.76%
	Jewish	1	4.76%
	No religious affiliation	1	4.76%
Christian	Christian	1	4.76%
	Evangelical, Jehovah's Witness, or Pentecostal	1	4.76%
Multiple	Buddhist or Hindu	2	9.52%
Disability		*13*	*100.00%*
Physical disability	Obese	4	30.77%
	Blindness or deafness	1	7.69%
	HIV infection	1	7.69%
	Spinal cord injury	1	7.69%
	Unspecified physical disability	1	7.69%
	Wheelchair user	1	7.69%
Mental disability	Asperger's Syndrome	1	7.69%
	Autism	1	7.69%
	Former depression	1	7.69%
	History of mental illness	1	7.69%
Age		*19*	*100.00%*
Older age	Age 37	2	10.53%
	Age 38	1	5.26%
	Age 44 or age 50	1	5.26%
	Age 46	1	5.26%
	Age 47	3	15.79%
	Age 47 or age 53	1	5.26%
	Age 50	1	5.26%
	Age 50 or age 51	1	5.26%
	Age 50, age 55, or age 62	1	5.26%
	Age 56	1	5.26%
	Age 57 or age 58	1	5.26%
	Age 60–61	1	5.26%
	Age 64–66	2	10.53%
Younger age	Age 22–23	1	5.26%
	Age 23	1	5.26%
Military service or affiliation		*4*	*100.00%*

Military service	Current membership in the Reserves	1	25.00%
	Military service	3	75.00%
Wealth		***7***	***100.00%***
Lower class	Residence in neighbourhood with bad reputation	3	42.86%
	Lower-class background	2	28.57%
	Non-upper-caste	2	28.57%
Marital status		***4***	***100.00%***
Married	Married	3	75.00%
	Married and childless	1	25.00%
Sexual orientation		***12***	***100.00%***
LGB+ organisation affiliation	LGB+ organisation affiliation	10	83.33%
Same-sex orientation	Same-sex marriage partner	1	8.33%
	Same-sex orientation	1	8.33%
Political orientation		***1***	***100.00%***
Political orientation	Orientation of mentioned youth political organisation	1	100.00%
Union affiliation		***1***	***100.00%***
Union affiliation	Mention of youth union membership	1	100.00%
Physical appearance		***9***	***100.00%***
Lower physical attractiveness	Lower physical attractiveness of resume picture	8	88.89%
Visible tattoo	Visible tattoo	1	11.11%

Notes. Abbreviations and notations used: LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations). Counts represent units of observation included in the meta-analysis.

Table A4. Main meta-analytic results by discrimination ground, adjusted for outliers

Variable		Effect			Statistical heterogeneity	
Discrimination ground	k	k-adj. $\widehat{DR}$ [$CI_{95\%}$]	t (p)	$\Delta_{k\text{-adj. }\widehat{DR}\text{ - }\widehat{DR}}$ (z, p)	I^2 [$CI_{95\%}$]	$\widehat{DR}^*$ $CI_{95\%}$
Race and national origin	98	0.6674 [0.6482; 0.6871]	−27.54*** (<0.001)	0.0074 (0.36, 0.715)	0.4421 [0.2903; 0.5613]	[0.5602; 0.7951]
Gender and motherhood status	51	1.0278 [0.9870; 1.0704]	1.36 (0.180)	0.0016 (0.04, 0.966)	0.6064 [0.4660; 0.7099]	[0.8292; 1.2740]
Age	17	0.6083 [0.5350; 0.6917]	−8.20*** (<0.001)	0.0092 (0.17, 0.866)	0.6490 [0.4123; 0.7904]	[0.3881; 0.9536]
Religion	13	0.7165 [0.6186; 0.8298]	−4.95*** (<0.001)	0.0246 (0.34, 0.732)	0.6739 [0.4172; 0.8175]	[0.4546; 1.1293]
Disability	12	0.7171 [0.6181; 0.8320]	−4.92*** (<0.001)	0.1580 (1.09, 0.276)	0.5621 [0.1645; 0.7705]	[0.4693; 1.0959]
Sexual orientation	9	0.9007 [0.7845; 1.0341]	−1.75 (0.119)	0.1991 (1.62, 0.104)	0.6900 [0.3800; 0.8450]	[0.6103; 1.3293]
Physical appearance	8	0.7195 [0.6196; 0.8354]	−5.21** (0.001)	0.0887 (0.94, 0.345)	0.8391 [0.6991; 0.9139]	[0.4844; 1.0686]
Wealth	7	0.8806 [0.8081; 0.9596]	−3.62* (0.011)	N/A	0.0000 [0.0000; 0.7081]	[0.8046; 0.9638]
Military service or affiliation	4	0.9983 [0.7766; 1.2834]	−0.02 (0.985)	N/A	0.6743 [0.0509; 0.8882]	[0.5513; 1.8080]
Marital status	4	0.8846 [0.8109; 0.9650]	−4.49* (0.021)	N/A	0.0000 [0.0000; 0.8469]	[0.7865; 0.9950]

Notes. Abbreviations and notations used: k (number of correspondence experiments), k-adj. $\widehat{DR}$ (pooled discrimination ratio estimate adjusted for outliers), $CI_{95\%}$ (95% confidence interval), $\Delta_{k\text{-adj. }\widehat{DR}\text{ - }\widehat{DR}}$ (the difference between the k-adjusted $\widehat{DR}$ and the non-adjusted $\widehat{DR}$), $\widehat{DR}^*$ $CI_{95\%}$ (95% prediction interval of the pooled discrimination ratio), LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations), and N/A (not applicable). $\Delta_{k\text{-adj. }\widehat{DR}\text{ - }\widehat{DR}}$ are only reported for differences greater than zero. Following Higgins and Thompson (2002), I^2 values around 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. A tabulated overview of the detected outliers can be retrieved from Table A11. * p < .05, ** p < .01, *** p < .001.

Table A5. Main meta-analytic results by discrimination ground and treatment group, adjusted for outliers

Variable		Effect			Statistical heterogeneity		Treatment group heterogeneity
Discrimination ground or treatment group	k	k-adj. $\widehat{DR}$ [$CI_{95\%}$]	t (p)	$\Delta_{k\text{-adj.}\ \widehat{DR}-\widehat{DR}}$ (z, p)	I^2 [$CI_{95\%}$]	$\widehat{DR}^*$ $CI_{95\%}$	Q (p)
Race and national origin	*130*	*0.6846 [0.6574; 0.7129]*	*−18.53*** (<0.001)*	*0.0246 (1.08, 0.278)*	*0.8505 [0.8270; 0.8708]*	*[0.4641; 1.0099]*	*59.22*** (<0.001)*
Arab/Maghrebi/Middle Eastern	27	0.5777 [0.5379; 0.6204]	−15.80*** (<0.001)	0.0380 (1.04, 0.299)	0.6440 [0.4638; 0.7636]	[0.4273; 0.7810]	
African/African American/Black	22	0.6391 [0.5874; 0.6952]	−11.05*** (<0.001)	−0.0011 (−0.02, 0.981)	0.5981 [0.3583; 0.7482]	[0.4846; 0.8427]	
Western Asian	15	0.7696 [0.7059; 0.8390]	−6.50*** (<0.001)	0.0472 (0.95, 0.344)	0.3458 [0.0000; 0.6473]	[0.6093; 0.9721]	
Eastern Asian/South-Eastern Asian	10	0.6381 [0.5288; 0.7699]	−5.41*** (<0.001)	0.0700 (0.74, 0.459)	0.8448 [0.7311; 0.9104]	[0.3562; 1.1430]	
Hispanic/Latin American/Caribbean	9	0.8175 [0.7095; 0.9420]	−3.28* (0.011)	−0.0535 (−0.61, 0.542)	0.4527 [0.0000; 0.7466]	[0.5844; 1.1435]	
Southern European	10	0.7027 [0.6168; 0.8005]	−6.12*** (<0.001)	N/A	0.7697 [0.5761; 0.8749]	[0.4935; 1.0005]	
Mixed/Multiple	7	0.8171 [0.6732; 0.9918]	−2.55* (0.043)	0.1414 (0.91, 0.361)	0.7784 [0.5398; 0.8933]	[0.5093; 1.3109]	N/A
Southern Asian/Indian	8	0.7004 [0.6352; 0.7723]	−8.61*** (<0.001)	N/A	0.5316 [0.0000; 0.7892]	[0.5642; 0.8695]	
Northern European/Western European	8	0.8154 [0.6661; 0.9981]	−2.39* (0.048)	N/A	0.7649 [0.5311; 0.8821]	[0.4738; 1.4031]	
Asian (generic)	5	0.6739 [0.4530; 1.0024]	−2.76 (0.051)	N/A	0.7777 [0.4645; 0.9077]	[0.2776; 1.6358]	
Eastern European	5	0.7206 [0.5271; 0.9851]	−2.91* (0.044)	N/A	0.9231 [0.8504; 0.9605]	[0.3122; 1.6635]	
Indigenous	3	0.7793 [0.4127; 1.4715]	−1.69 (0.233)	N/A	0.9571 [0.9065; 0.9803]	[0.0199; 30.5504]	
Central Asian	1	N/A	N/A	N/A	N/A	N/A	
Gender and motherhood status	*54*	*1.0401 [0.9963; 1.0857]*	*1.83 (0.072)*	*0.0138 (0.36, 0.718)*	*0.6595 [0.5470; 0.7441]*	*[0.8152; 1.3269]*	*31.55*** (<0.001)*
Female gender	44	1.0663 [1.0221; 1.1124]	3.06** (0.004)	0.0181 (0.44, 0.659)	0.6413 [0.5057; 0.7398]	[0.8614; 1.3199]	
Mother	8	0.9044 [0.7887; 1.0370]	−1.74 (0.126)	N/A	0.3049 [0.0000; 0.6902]	[0.6806; 1.2017]	N/A
Transgender	2	0.8500 [0.5306; 1.3619]	−4.38 (0.143)	N/A	N/A	N/A	
Age	*17*	*0.6083 [0.535; 0.6917]*	*−8.20*** (<0.001)*	*0.0092 (0.17, 0.866)*	*0.649 [0.4123; 0.7904]*	*[0.3881; 0.9536]*	*5.46* (0.019)*
Older age	15	0.5894 [0.5147; 0.6748]	−8.38*** (<0.001)	0.0089 (0.16, 0.873)	0.6643 [0.422; 0.8051]	[0.3786; 0.9175]	N/A
Younger age	2	0.7698 [0.2294; 2.5830]	−2.75 (0.222)	N/A	N/A	N/A	
Religion	*20*	*0.7151 [0.6168; 0.8291]*	*−4.75*** (<0.001)*	*0.0232 (0.32, 0.751)*	*0.9237 [0.8960; 0.9441]*	*[0.3838; 1.3324]*	*15.37** (0.002)*
Muslim	13	0.6653 [0.5526; 0.8010]	−4.78*** (<0.001)	0.0304 (0.37, 0.713)	0.8476 [0.7547; 0.9053]	[0.3547; 1.2479]	
Other	3	0.8240 [0.3578; 1.8979]	−1.00 (0.423)	N/A	0.9132 [0.7762; 0.9663]	[0.0066; 103.6074]	N/A
Christian	2	0.7293 [0.0075; 71.1483]	−0.88 (0.542)	N/A	N/A	N/A	
Multiple	2	0.9275 [0.7532; 1.1422]	−4.59 (0.137)	N/A	N/A	N/A	
Disability	*12*	*0.7171 [0.6181; 0.8320]*	*−4.92*** (<0.001)*	*0.1580 (1.09, 0.276)*	*0.5621 [0.1645; 0.7705]*	*[0.4693; 1.0959]*	*1.39 (0.239)*
Physical disability	8	0.7494 [0.6259; 0.8974]	−3.78** (0.007)	0.2125 (1.03, 0.301)	0.5680 [0.0498; 0.8036]	[0.4704; 1.1939]	N/A
Mental disability	4	0.6249 [0.4075; 0.9581]	−3.50* (0.039)	N/A	0.3843 [0.0000; 0.7895]	[0.2498; 1.5630]	
Sexual orientation	*10*	*0.8358 [0.6804; 1.0267]*	*−1.97 (0.080)*	*0.1342 (1.04, 0.298)*	*0.9354 [0.9010; 0.9579]*	*[0.4318; 1.6179]*	*6.25* (0.012)*

LGB+ organisation affiliation	8	0.7924 [0.6203; 1.0122]	−2.25 (0.059)	0.1441 (1.07, 0.287)	0.9482 [0.9185; 0.9670]	[0.3816; 1.6454]	N/A
Same-sex orientation	2	1.0585 [0.5470; 2.0485]	1.09 (0.471)	N/A	N/A	N/A	

Notes. Abbreviations and notations used: *k* (number of correspondence experiments), *k*-adj. $\widehat{DR}$ (pooled discrimination ratio estimate adjusted for outliers), $CI_{95\%}$ (95% confidence interval), $\Delta_{k\text{-adj.}\,\widehat{DR}\,-\,\widehat{DR}}$ (the difference between the *k*-adjusted $\widehat{DR}$ and the non-adjusted $\widehat{DR}$), $\widehat{DR}^{*}$ $CI_{95\%}$ (95% prediction interval of the pooled discrimination ratio), LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations), and N/A (not applicable). Following Schwarzer et al. (2015), treatment group heterogeneity statistics are only given for discrimination grounds for which $k \geq 10$ (see Table A4 for the results related to physical appearance, wealth, military service or affiliation, and marital status). Pooled discrimination rates are only calculated for the discrimination grounds or treatment groups for which $k > 1$, while statistical heterogeneity statistics are only calculated for those grounds or groups for which $k > 2$. $\Delta_{k\text{-adj.}\,\widehat{DR}\,-\,\widehat{DR}}$ are only reported for differences greater than zero. Following Higgins and Thompson (2002), I^2 values around 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. A tabulated overview of the detected outliers can be retrieved from Table A11. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A6. Robustness analyses: bias statistics and 'limit' meta-analyses

Variable	Sample	Bias statistic	Limit meta-analysis		
Discrimination ground or treatment group	k	β (t, p)	limit-adj. $\widehat{DR}$ [$CI_{95\%}$]	z (p)	$\Delta_{\text{limit-adj. } \widehat{DR} - \widehat{DR}}$ (z, p)
Race and national origin	***143***	***−3.64 (−0.09, 0.927)***	***0.7113 [0.6924; 0.7307]***	***−24.79*** (<0.001)***	***0.0513 (2.482, 0.013)***
Arab/Maghrebi/Middle Eastern	31	36.28 (0.30, 0.767)	0.5937 [0.5548; 0.6353]	−15.09*** (<0.001)	0.0540 (1.46, 0.144)
African/African American/Black	26	−71.57 (−0.83, 0.417)	0.6845 [0.6444; 0.7270]	−12.32*** (<0.001)	0.0443 (1.03, 0.301)
Western Asian	17	105.56 (1.85, 0.084)	0.7508 [0.6977; 0.8080]	−7.66*** (<0.001)	0.0284 (0.59, 0.554)
Eastern Asian/South-Eastern Asian	11	328.63 (0.68, 0.512)	0.6286 [0.5368; 0.7361]	−5.77*** (<0.001)	0.0605 (0.65, 0.515)
Hispanic/Latin American/Caribbean	10	−4.62 (−0.05, 0.959)	0.9220 [0.8091; 1.0507]	−1.22 (0.223)	0.0510 (0.53, 0.595)
Southern European	10	84.95 (0.79, 0.454)	0.6673 [0.5711; 0.7798]	−5.09*** (<0.001)	−0.0354 (−0.53, 0.598)
Gender and motherhood status	***72***	***66.95 (0.78, 0.437)***	***1.0413 [1.0151; 1.0682]***	***3.11** (0.002)***	***0.0151 (0.44, 0.657)***
Female gender	62	123.72 (1.09, 0.279)	1.0413 [1.0138; 1.0696]	2.97** (0.003)	−0.0068 (−0.18, 0.854)
Age	***19***	***−100.35 (−0.86, 0.401)***	***0.6867 [0.6503; 0.7250]***	***−13.54*** (<0.001)***	***0.0876 (1.88, 0.060)***
Older age	17	−138.13 (−1.14, 0.271)	0.6646 [0.6292; 0.7020]	−14.64*** (<0.001)	0.0841 (1.77, 0.076)
Religion	***21***	***−191.6 (−1.19, 0.249)***	***0.7855 [0.7457; 0.8274]***	***−9.11*** (<0.001)***	***0.0936 (1.568, 0.117)***
Muslim	14	−176.8 (−1.23, 0.242)	0.7730 [0.7069; 0.8452]	−5.65*** (<0.001)	0.138 (1.88, 0.060)
Disability	***13***	***586.83 (0.98, 0.347)***	***0.5885 [0.5277; 0.6563]***	***−9.53*** (<0.001)***	***0.0293 (0.23, 0.820)***
Sexual orientation	***12***	***−117.28 (−0.14, 0.890)***	***0.5287 [0.5101; 0.5480]***	***−34.84*** (<0.001)***	***−0.1729* (−1.98, 0.047)***
LGB+ organisation affiliation	10	−513.99 (−0.49, 0.634)	0.5545 [0.5343; 0.5755]	−31.19*** (<0.001)	−0.0937 (−0.98, 0.325)

Notes. Abbreviations and notations used: k (number of correspondence experiments), limit-adj. $\widehat{DR}$ (pooled discrimination ratio estimate based on the 'limit' meta-analysis), $CI_{95\%}$ (95% confidence interval), $\Delta_{\text{limit-adj. } \widehat{DR} - \widehat{DR}}$ (the difference between the limit-adjusted $\widehat{DR}$ and the non-adjusted $\widehat{DR}$), and LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations). Bias statistic (β) is derived from the binary-effects adaptation of Egger's regression test by Peter et al. (2006)—the null hypothesis assumes that there is no funnel plot asymmetry. 'Limit' meta-analyses produce more conservative but more precise 'shrunken' estimates of the pooled discrimination ratios (limit-adj. $\widehat{DR}$; Schwarzer et al., 2020). Following Harrer et al. (2021), statistics are only given for discrimination grounds or treatment groups for which $k \geq 10$. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A7. Sub-group heterogeneity of the pooled discrimination ratios by discrimination ground, adjusted for outliers

Variable				Effect			Statistical heterogeneity		Sub-group heterogeneity
Discrimination ground	Sub-group	Level	*k*	*k*-adj. $\widehat{DR}$ [$CI_{95\%}$]	*t* (*p*)	$\Delta_{k\text{-adj.}\ \widehat{DR}\ -\ \widehat{DR}}$ (z, p)	I^2 [$CI_{95\%}$]	$\widehat{DR}^{*}$ $CI_{95\%}$	*Q* (*p*)
Race and national origin	Call-back	Stricto	65	0.6668 [0.6426; 0.6919]	−21.91*** (<0.001)	0.0076 (0.31, 0.755)	0.4181 [0.2161; 0.5680]	[0.5600; 0.7938]	0.01 (0.918)
		Lato	33	0.6689 [0.6363; 0.7031]	−16.43*** (<0.001)	0.0089 (0.25, 0.803)	0.4963 [0.2463; 0.6634]	[0.5500; 0.8134]	
	Region	Americas	29	0.6861 [0.6484; 0.7261]	−13.64*** (<0.001)	−0.0138 (−0.31, 0.756)	0.1980 [0.0000; 0.4937]	[0.5922; 0.7949]	2.92 (0.404)
		Europe	62	0.6664 [0.6423; 0.6913]	−22.06*** (<0.001)	0.0057 (0.25, 0.803)	0.4979 [0.3260; 0.6259]	[0.5528; 0.8032]	
		Asia	4	0.6049 [0.4808; 0.7612]	−6.96** (0.006)	0.1172 (1.00, 0.316)	0.3417 [0.0000; 0.7688]	[0.3627; 1.0091]	
		Other	3	0.6555 [0.4950; 0.8681]	−6.47* (0.023)	0.0280 (0.26, 0.792)	0.5243 [0.0000; 0.8633]	[0.1747; 2.4600]	
	Period	2002–2010	32	0.6629 [0.6317; 0.6957]	−17.40*** (<0.001)	0.0153 (0.55, 0.582)	0.3414 [0.0000; 0.5727]	[0.5695; 0.7717]	0.10 (0.749)
		2011–2020	66	0.6693 [0.6450; 0.6946]	−21.61*** (<0.001)	0.0014 (0.05, 0.959)	0.4842 [0.3127; 0.6130]	[0.5513; 0.8126]	
Gender and motherhood status	Call-back	Stricto	38	1.0253 [0.9824; 1.0701]	1.19 (0.243)	−0.0135 (−0.32, 0.746)	0.5795 [0.3976; 0.7065]	[0.8537; 1.2314]	0.16 (0.685)
		Lato	13	1.0490 [0.9365; 1.1749]	0.92 (0.377)	0.0482 (0.58, 0.559)	0.6924 [0.4548; 0.8265]	[0.7269; 1.5136]	
	Region	Americas	5	0.9587 [0.5990; 1.5345]	−0.25 (0.816)	−0.0071 (−0.04, 0.972)	0.5552 [0.0000; 0.8356]	[0.3146; 2.9216]	7.60 (0.055)
		Europe	35	1.0224 [0.9724; 1.0750]	0.90 (0.376)	0.0083 (0.18, 0.854)	0.5451 [0.3334; 0.6895]	[0.8214; 1.2726]	
		Asia	10	1.0133 [0.9333; 1.1001]	0.36 (0.726)	−0.0338 (−0.56, 0.572)	0.7012 [0.4272; 0.8441]	[0.8182; 1.2548]	
		Other	1	1.1804 [1.0699; 1.3022]	3.31*** (<0.001)	−0.1051 (−0.76, 0.446)	N/A	N/A	
	Period	2002–2010	17	1.0670 [0.9967; 1.1422]	2.02 (0.061)	−0.0566 (−1.00, 0.319)	0.4736 [0.0745; 0.7006]	[0.8620; 1.3207]	1.99 (0.159)
		2011–2020	34	1.0073 [0.9571; 1.0602]	0.29 (0.773)	0.0248 (0.54, 0.586)	0.6409 [0.4825; 0.7508]	[0.8070; 1.2574]	
Age	Call-back	Stricto	12	0.5941 [0.4898; 0.7206]	−5.94*** (<0.001)	0.0067 (0.09, 0.929)	0.6175 [0.2835; 0.7958]	[0.3328; 1.0605]	0.19 (0.667)
		Lato	5	0.6248 [0.5038; 0.7749]	−6.07** (0.004)	N/A	0.7404 [0.3551; 0.8955]	[0.3624; 1.0771]	
	Region	Americas	6	0.6881 [0.6438; 0.7354]	−14.44*** (<0.001)	N/A	0.2992 [0.0000; 0.7140]	[0.6195; 0.7642]	9.01** (0.003)
		Europe	11	0.5347 [0.4475; 0.6390]	−7.83*** (<0.001)	−0.0020 (−0.03, 0.976)	0.5392 [0.0893; 0.7668]	[0.3323; 0.8603]	
		Asia	N/A	N/A	N/A	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	6	0.546 [0.4149; 0.7185]	−5.67** (0.002)	N/A	0.5205 [0.0000; 0.8088]	[0.2967; 1.0047]	1.50 (0.220)
		2011–2020	11	0.639 [0.5452; 0.7488]	−6.29*** (<0.001)	0.0141 (0.20, 0.838)	0.6661 [0.3694; 0.8232]	[0.3976; 1.0270]	
Religion	Call-back	Stricto	8	0.7327 [0.5755; 0.9329]	−3.05* (0.019)	0.0533 (0.49, 0.626)	0.6943 [0.3628; 0.8534]	[0.3802; 1.4122]	0.21 (0.649)
		Lato	5	0.6885 [0.5348; 0.8864]	−4.10* (0.015)	−0.0135 (−0.14, 0.889)	0.7118 [0.2698; 0.8863]	[0.3935; 1.2048]	
	Region	Americas	4	0.6745 [0.4376; 1.0395]	−2.90 (0.063)	−0.0791 (−0.57, 0.570)	0.6723 [0.0443; 0.8876]	[0.2354; 1.9324]	3.24 (0.197)
		Europe	8	0.7114 [0.5868; 0.8625]	−4.18** (0.004)	0.0282 (0.33, 0.744)	0.6932 [0.3599; 0.8529]	[0.4237; 1.1944]	
		Asia	1	1.0293 [0.6885; 1.5386]	0.14 (0.888)	0.3956 (0.94, 0.349)	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	3	0.5797 [0.1388; 2.4207]	−1.64 (0.242)	0.0164 (0.07, 0.943)	0.8257 [0.4660; 0.9431]	[0.0002; >1000.0000]	0.61 (0.434)
		2011–2020	10	0.7537 [0.6782; 0.8375]	−6.06*** (<0.001)	0.0305 (0.45, 0.653)	0.5188 [0.0121; 0.7656]	[0.5902; 0.9624]	

Disability	Call-back	Stricto	10	0.7066 [0.5837; 0.8555]	−4.11** (0.003)	0.1735 (1.04, 0.298)	0.6323 [0.2725; 0.8142]	[0.4192; 1.1912]	0.60 (0.438)
		Lato	2	0.7586 [0.4853; 1.1859]	−7.86 (0.081)	N/A	N/A	N/A	
	Region	Americas	5	0.7308 [0.5198; 1.0275]	−2.56 (0.063)	N/A	0.7041 [0.2466; 0.8838]	[0.3245; 1.6458]	0.04 (0.847)
		Europe	7	0.7102 [0.5786; 0.8719]	−4.08** (0.006)	0.2380 (1.16, 0.245)	0.4796 [0.0000; 0.7802]	[0.4400; 1.1465]	
		Asia	N/A	N/A	N/A	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	2	0.8138 [0.7514; 0.8814]	−32.80* (0.019)	0.4979 (1.00, 0.318)	N/A	N/A	4.30* (0.038)
		2011–2020	10	0.6863 [0.5701; 0.8261]	−4.59** (0.001)	N/A	0.6176 [0.2390; 0.8078]	[0.4217; 1.1167]	
Sexual orientation	Call-back	Stricto	7	0.8717 [0.7294; 1.0417]	−1.89 (0.108)	0.2182 (1.63, 0.103)	0.7162 [0.3841; 0.8692]	[0.5480; 1.3866]	1.54 (0.215)
		Lato	2	0.9970 [0.3606; 2.7565]	−0.04 (0.976)	N/A	N/A	N/A	
	Region	Americas	3	0.7650 [0.4181; 1.4000]	−1.91 (0.197)	N/A	0.7581 [0.2036; 0.9265]	[0.0304; 19.2721]	2.05 (0.153)
		Europe	6	0.9427 [0.8511; 1.0442]	−1.48 (0.198)	0.1692 (1.25, 0.213)	0.4674 [0.0000; 0.7891]	[0.7573; 1.1735]	
		Asia	N/A	N/A	N/A	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	3	0.8365 [0.4274; 1.6370]	−1.14 (0.371)	0.2206 (1.18, 0.237)	0.8806 [0.6667; 0.9573]	[0.0183; 38.3324]	0.42 (0.516)
		2011–2020	6	0.9298 [0.8237; 1.0496]	−1.54 (0.183)	0.1580 (0.92, 0.359)	0.2733 [0.0000; 0.6991]	[0.7275; 1.1883]	

Notes. Abbreviations and notations used: *k* (number of correspondence experiments), $\widehat{DR}$ (pooled discrimination ratio estimate), $CI_{95\%}$ (95% confidence interval), $\Delta_{k\text{-adj. }\widehat{DR}\text{ - }\widehat{DR}}$ (the difference between the *k*-adjusted $\widehat{DR}$ and the non-adjusted $\widehat{DR}$), $\widehat{DR}^{*}$ $CI_{95\%}$ (95% prediction interval of the pooled discrimination ratio), and N/A (not applicable). 'Stricto' refers to correspondence experiments in which the call-back variable is related to an invitation to a job interview, while 'Lato' refers to experiments in which said variable concerns any positive reaction to an application (e.g. a request of the employer for additional information about the applicant). Following Harrer et al. (2021) and Schwarzer et al. (2015), sub-group heterogeneity statistics are only calculated for the discrimination grounds for which $k \geq 10$, while statistical heterogeneity statistics are only calculated for those groups for which $k > 2$. $\Delta_{k\text{-adj. }\widehat{DR}\text{ - }\widehat{DR}}$ are only reported for differences greater than zero. Following Higgins and Thompson (2002), I^2 values around 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. A tabulated overview of the detected outliers can be retrieved from Table A11. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A8. Sub-group heterogeneity of the pooled discrimination ratios by treatment group, adjusted for outliers

Variable			Effect				*Statistical heterogeneity*		*Sub-group heterogeneity*
Treatment group	Sub-group	Level	*k*	*k-adj.* $\widehat{DR}$ [$CI_{95\%}$]	*t* (*p*)	$\Delta_{k\text{-adj. }\widehat{DR}-\widehat{DR}}$ (*z*, *p*)	I^2 [$CI_{95\%}$]	$\widehat{DR}^*$ $CI_{95\%}$	*Q* (*p*)
Race and national origin									
Arab/Maghrebi/Middle Eastern	Call-back	Stricto	16	0.5889 [0.5544; 0.6255]	−18.72*** (<0.001)	0.0201 (0.51, 0.608)	0.5296 [0.1695; 0.7335]	[0.5008; 0.6924]	1.10 (0.295)
		Lato	11	0.5343 [0.4387; 0.6507]	−7.08*** (<0.001)	0.0406 (0.58, 0.563)	0.7551 [0.5584; 0.8642]	[0.2884; 0.9900]	
	Region	Americas	2	0.5967 [0.0430; 8.2789]	−2.49 (0.243)	N/A	N/A	N/A	1.14 (0.566)
		Europe	24	0.5724 [0.5290; 0.6193]	−14.64*** (<0.001)	0.0385 (0.98, 0.327)	0.6722 [0.4983; 0.7858]	[0.4143; 0.7908]	
		Asia	N/A	N/A	N/A	N/A	N/A	N/A	
		Other	1	0.6288 [0.5375; 0.7356]	−5.79*** (<0.001)	N/A	N/A	N/A	
	Period	2002–2010	11	0.5654 [0.5093; 0.6276]	−12.16*** (<0.001)	0.0067 (0.13, 0.898)	0.7014 [0.4454; 0.8393]	[0.4161; 0.7682]	0.32 (0.574)
		2011–2020	16	0.5878 [0.5274; 0.6551]	−10.45*** (<0.001)	0.0595 (1.16, 0.246)	0.5830 [0.2746; 0.7603]	[0.4105; 0.8416]	
African American/Black	Call-back	Stricto	14	0.6534 [0.5736; 0.7444]	−7.05*** (<0.001)	−0.0190 (−0.34, 0.736)	0.4894 [0.0539; 0.7245]	[0.4601; 0.9280]	0.53 (0.467)
		Lato	8	0.6160 [0.5423; 0.6999]	−8.98*** (<0.001)	−0.0024 (−0.03, 0.973)	0.5632 [0.0380; 0.8017]	[0.4595; 0.8260]	
	Region	Americas	14	0.6771 [0.6128; 0.7481]	−8.45*** (<0.001)	0.0327 (0.54, 0.586)	0.3335 [0.0000; 0.6480]	[0.5335; 0.8592]	9.53** (0.009)
		Europe	7	0.5606 [0.4921; 0.6387]	−10.86*** (<0.001)	−0.0550 (−0.95, 0.343)	0.2659 [0.0000; 0.6812]	[0.4406; 0.7132]	
		Asia	1	0.8000 [0.5874; 1.0895]	−1.42 (0.157)	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	7	0.6162 [0.5268; 0.7208]	−7.56*** (<0.001)	N/A	0.3550 [0.0000; 0.7277]	[0.4651; 0.8164]	0.033 (0.568)
		2011–2020	15	0.6460 [0.5775; 0.7227]	−8.35*** (<0.001)	0.0038 (0.06, 0.948)	0.6591 [0.4118; 0.8024]	[0.4621; 0.9031]	
Western Asian	Call-back	Stricto	12	0.7408 [0.656; 0.8366]	−5.43*** (<0.001)	0.0492 (0.81, 0.416)	0.3588 [0.0000; 0.6762]	[0.5502; 0.9975]	2.11 (0.147)
		Lato	3	0.8078 [0.7332; 0.8900]	−9.48* (0.011)	N/A	0.0000 [0.0000; 0.8960]	[0.6069; 1.0753]	
	Region	Americas	N/A	N/A	N/A	N/A	N/A	N/A	1.86 (0.172)
		Europe	14	0.7772 [0.7138; 0.8463]	−6.40*** (<0.001)	N/A	0.3289 [0.0000; 0.6456]	[0.6229; 0.9699]	
		Asia	1	0.5909 [0.4017; 0.8692]	−2.67** (0.008)	0.0588 (0.51, 0.607)	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	3	0.7795 [0.4617; 1.3161]	−2.05 (0.177)	0.0912 (0.64, 0.520)	0.7120 [0.0216; 0.9152]	[0.0463; 13.1326]	0.02 (0.883)
		2011–2020	12	0.7650 [0.7028; 0.8326]	−6.96*** (<0.001)	0.0336 (0.68, 0.494)	0.1902 [0.0000; 0.5800]	[0.6396; 0.9150]	
Eastern Asian/South-Eastern Asian	Call-back	Stricto	6	0.5966 [0.4457; 0.7986]	−4.55** (0.006)	0.0960 (0.80, 0.425)	0.8563 [0.7073; 0.9295]	[0.2773; 1.2834]	0.91 (0.339)
		Lato	4	0.6996 [0.4744; 1.0317]	−2.93 (0.061)	N/A	0.8624 [0.6648; 0.9435]	[0.2335; 2.0958]	
	Region	Americas	2	0.7824 [0.5385; 1.1369]	−8.34 (0.076)	N/A	N/A	N/A	20.34*** (<0.001)
		Europe	5	0.6759 [0.5079; 0.8996]	−3.80* (0.019)	N/A	0.8401 [0.6401; 0.929]	[0.3294; 1.3871]	
		Asia	1	0.5941 [0.5269; 0.6699]	−8.50*** (<0.001)	0.2568 (0.98, 0.325)	N/A	N/A	
		Other	2	0.4689 [0.0182; 12.0734]	−2.96 (0.207)	N/A	N/A	N/A	
	Period	2002–2010	4	0.6886 [0.5218; 0.9085]	−4.28* (0.023)	N/A	0.5545 [0.0000; 0.8525]	[0.3604; 1.3156]	0.61 (0.437)
		2011–2020	6	0.6108 [0.4407; 0.8466]	−3.88* (0.012)	0.0949 (0.72, 0.472)	0.9019 [0.8136; 0.9483]	[0.2487; 1.5004]	
Southern European	Call-back	Stricto	7	0.7234 [0.6156; 0.8501]	−4.91** (0.003)	N/A	0.7851 [0.5563; 0.8959]	[0.4898; 1.0685]	0.47 (0.492)
		Lato	3	0.6561 [0.3818; 1.1273]	−3.35 (0.079)	N/A	0.7169 [0.0407; 0.9164]	[0.0419; 10.2836]	
	Region	Americas	2	0.7615 [0.3710; 1.5629]	−4.82 (0.130)	N/A	N/A	N/A	12.87** (0.002)
		Europe	7	0.6553 [0.5635; 0.7621]	−6.85*** (<0.001)	N/A	0.6464 [0.2033; 0.8430]	[0.4618; 0.9299]	
		Asia	N/A	N/A	N/A	N/A	N/A	N/A	

		Other	1	0.9134 [0.7978; 1.0458]	−1.31 (0.190)	N/A	N/A	N/A	
	Period	2002–2010	5	0.7189 [0.5609; 0.9213]	−3.69* (0.021)	N/A	0.8544 [0.6784; 0.9341]	[0.3866; 1.3366]	0.20 (0.655)
		2011–2020	5	0.6814 [0.5455; 0.8510]	−4.79** (0.009)	N/A	0.6127 [0.0000; 0.8545]	[0.4282; 1.0842]	
Gender and motherhood status									
Female gender	Call-back	Stricto	31	1.0524 [1.0079; 1.0989]	2.41* (0.022)	−0.0056 (−0.12, 0.901)	0.5649 [0.3500; 0.7088]	[0.8923; 1.2413]	0.86 (0.354)
		Lato	13	1.1042 [0.9947; 1.2259]	2.07 (0.061)	0.0750 (0.83, 0.405)	0.7564 [0.5822; 0.8580]	[0.7842; 1.5550]	
	Region	Americas	6	1.1641 [1.0392; 1.3041]	3.44* (0.018)	0.1573 (1.21, 0.226)	0.5318 [0.0000; 0.8129]	[0.9216; 1.4705]	6.54 (0.088)
		Europe	27	1.0460 [0.9892; 1.1061]	1.66 (0.110)	0.0066 (0.12, 0.901)	0.5810 [0.3582; 0.7264]	[0.8380; 1.3057]	
		Asia	9	1.0365 [0.9619; 1.1168]	1.11 (0.301)	−0.0106 (−0.18, 0.855)	0.6154 [0.2045; 0.8141]	[0.8727; 1.2311]	
		Other	2	1.3965 [0.0708; 27.536]	1.42 (0.390)	0.1111 (0.32, 0.745)	N/A	N/A	
	Period	2002–2010	17	1.1082 [1.0488; 1.1709]	3.96** (0.001)	−0.036 (−0.70, 0.485)	0.4394 [0.008; 0.6832]	[0.9454; 1.2991]	2.68 (0.102)
		2011–2020	27	1.0383 [0.9762; 1.1044]	1.25 (0.221)	0.0399 (0.73, 0.464)	0.6684 [0.5044; 0.7782]	[0.8070; 1.3360]	
Age									
Older age	Call-back	Stricto	10	0.5589 [0.4508; 0.6930]	−6.12*** (<0.001)	0.0017 (0.02, 0.983)	0.5954 [0.1883; 0.7984]	[0.3123; 1.0005]	0.82 (0.364)
		Lato	5	0.6248 [0.5038; 0.7749]	−6.07** (0.004)	N/A	0.7404 [0.3551; 0.8955]	[0.3624; 1.0771]	
	Region	Americas	5	0.6916 [0.6342; 0.7541]	−11.83*** (<0.001)	N/A	0.4361 [0.0000; 0.7929]	[0.5877; 0.8138]	19.18*** (<0.001)
		Europe	10	0.5221 [0.4599; 0.5927]	−11.59*** (<0.001)	0.0069 (0.13, 0.897)	0.1728 [0.0000; 0.5836]	[0.4199; 0.6492]	
		Asia	N/A	N/A	N/A	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	6	0.5460 [0.4149; 0.7185]	−5.67** (0.002)	N/A	0.5205 [0.0000; 0.8088]	[0.2967; 1.0047]	0.75 (0.387)
		2011–2020	9	0.6128 [0.5099; 0.7364]	−6.14*** (<0.001)	0.0118 (0.16, 0.874)	0.6992 [0.4012; 0.8489]	[0.3675; 1.0217]	
Religion									
Muslim	Call-back	Stricto	8	0.7327 [0.5755; 0.9329]	−3.05* (0.019)	0.0644 (0.55, 0.581)	0.6943 [0.3628; 0.8534]	[0.3802; 1.4122]	1.96 (0.161)
		Lato	5	0.5795 [0.4009; 0.8379]	−4.11* (0.015)	N/A	0.9253 [0.8556; 0.9614]	[0.2231; 1.5054]	
	Region	Americas	2	0.7019 [0.0303; 16.2621]	−1.43 (0.388)	N/A	N/A	N/A	0.04 (0.978)
		Europe	9	0.6723 [0.5500; 0.8216]	−4.56** (0.002)	0.0441 (0.49, 0.622)	0.8572 [0.7479; 0.9191]	[0.3747; 1.206]	
		Asia	2	0.6336 [0.0015; 265.6794]	−0.96 (0.513)	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	2	0.5840 [0.0003; >1000.0000]	−0.91 (0.531)	N/A	N/A	N/A	0.05 (0.817)
		2011–2020	11	0.6704 [0.5659; 0.7943]	−5.25*** (<0.001)	0.0320 (0.42, 0.676)	0.8554 [0.7587; 0.9133]	[0.3934; 1.1425]	
Sexual orientation									
LGB+ organisation affiliation	Call-back	Stricto	6	0.7324 [0.5359; 1.0009]	−2.56 (0.050)	0.1509 (1.08, 0.282)	0.9610 [0.9364; 0.9761]	[0.3103; 1.7286]	4.49* (0.034)
		Lato	2	0.9970 [0.3606; 2.7565]	−0.04 (0.976)	N/A	N/A	N/A	
	Region	Americas	2	0.6654 [0.2679; 1.6523]	−5.69 (0.111)	N/A	N/A	N/A	2.29 (0.130)
		Europe	6	0.8337 [0.5958; 1.1665]	−1.39 (0.223)	0.0974 (0.59, 0.558)	0.9590 [0.9327; 0.9751]	[0.3264; 2.1295]	
		Asia	N/A	N/A	N/A	N/A	N/A	N/A	
		Other	N/A	N/A	N/A	N/A	N/A	N/A	
	Period	2002–2010	4	0.7114 [0.3852; 1.3136]	−1.77 (0.175)	0.0955 (0.51, 0.610)	0.9668 [0.9398; 0.9817]	[0.1149; 4.4025]	1.42 (0.234)
		2011–2020	4	0.9012 [0.7716; 1.0525]	−2.13 (0.123)	0.2181 (1.05, 0.294)	0.2447 [0.0000; 0.8843]	[0.6326; 1.2836]	

Notes. Abbreviations and notations used: k (number of correspondence experiments), $\widehat{DR}$ (pooled discrimination ratio estimate), $CI_{95\%}$ (95% confidence interval), $\Delta_{k\text{-adj.}\ \widehat{DR}\ -\ \widehat{DR}}$ (the difference between the k-adjusted $\widehat{DR}$ and the non-adjusted $\widehat{DR}$), $\widehat{DR}^{*}$ $CI_{95\%}$ (95% prediction interval of the pooled discrimination ratio), LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations), and N/A (not applicable). 'Stricto' refers to correspondence experiments in which the call-back variable is related to an invitation to a job interview, while 'Lato' refers to experiments in which said variable concerns any positive reaction to an application (e.g. a request of the employer for additional information about the applicant). Following Harrer et al. (2021) and Schwarzer et al. (2015), sub-group heterogeneity statistics are only calculated for the treatment groups for which $k \geq 10$, while statistical heterogeneity statistics are only calculated for those groups for which $k > 2$. $\Delta_{k\text{-adj.}\ \widehat{DR}\ -\ \widehat{DR}}$ are only reported for differences greater than zero. Following Higgins and Thompson (2002), I^2 values around 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. A tabulated overview of the detected outliers can be retrieved from Table A11. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A9. Three-level mixed models to evaluate statistical dependencies between the pooled discrimination ratios

Discrimination ground or treatment group	Model (*df*)	AIC	AICc	BIC	Log-likelihood	Likelihood ratio test (*p*)
Race and national origin						
Arab/Maghrebi/Middle Eastern	Full (3)	19.94	20.86	24.14	−6.97	1.71 (0.191)
	Reduced (2)	19.65	20.09	22.45	−7.82	
Hispanic/Latin American/Caribbean	Full (3)	7.68	12.48	8.27	−0.84	0.18 (0.668)
	Reduced (2)	5.86	7.86	6.26	−0.93	
Mixed/Multiple	Full (3)	16.34	24.34	16.18	−5.17	1.09 (0.297)
	Reduced (2)	15.43	18.43	15.32	−5.71	
Gender and motherhood status						
Female gender	Full (3)	14.46	14.88	20.79	−4.23	1.18 (0.277)
	Reduced (2)	13.64	13.85	17.86	−4.82	
Age						
Older age	Full (3)	12.55	14.55	14.87	−3.27	0.32 (0.569)
	Reduced (2)	10.87	11.80	12.42	−3.44	
Religion						
Muslim	Full (3)	14.93	17.60	16.63	−4.47	0.96 (0.326)
	Reduced (2)	13.90	15.10	15.03	−4.95	

Notes. Abbreviations and notations used: *df* (degrees of freedom), AIC (Akaike Information Criterion), AICc (Akaike Information Criterion corrected for small sample sizes), and BIC (Bayesian Information Criterion). 'Full' three-level models include estimates of between- and within-study heterogeneity, while 'reduced' two-level models only include estimates of within-study heterogeneity. Statistics are only reported if the full model is different from the reduced model (i.e. model comparisons for which the likelihood ratio test statistic differs from zero). These results indicate that there are no statistical dependencies between the pooled discrimination ratios.

Table A10. Correlation statistics of the relationship between the research period and the weighted majority/minority response ratios

Discrimination ground, treatment group, or region	*k*	$\hat{r}_{Pearson}$	β	*t* (*p*)
Race and national origin	143	−0.25	−0.0117	−3.05** (0.003)
Arab/Maghrebi/Middle Eastern	31	−0.06	−0.0019	−0.33 (0.747)
African American/Black	26	−0.04	−0.0018	−0.20 (0.843)
Western Asian	17	−0.30	−0.0124	−1.22 (0.243)
Eastern Asian/South-Eastern Asian	11	−0.25	−0.0137	−0.76 (0.468)
Southern European	10	0.08	0.0027	0.24 (0.818)
Americas	38	0.05	0.0033	0.31 (0.755)
Europe	94	−0.40	−0.0164	−4.13*** (<0.001)
Gender and motherhood status	72	0.16	0.0098	1.34 (0.184)
Female gender	62	0.15	0.0099	1.21 (0.230)
Americas	10	0.08	0.0069	0.24 (0.818)
Europe	48	0.15	0.0094	1.05 (0.297)
Asia	11	0.30	0.0117	0.94 (0.372)
Age	19	−0.31	−0.0105	−1.36 (0.193)
Older age	17	−0.31	−0.0101	−1.26 (0.226)
Europe	13	−0.20	−0.0080	−0.66 (0.520)
Religion	21	−0.13	−0.0078	−0.55 (0.587)
Muslim	14	0.30	0.0169	1.11 (0.290)
Europe	14	−0.50	−0.0365	−2.01 (0.067)
Disability	13	−0.17	−0.0098	−0.58 (0.572)
Sexual orientation	12	−0.54	−0.0574	−2.02 (0.071)
LGB+ organisation affiliation	10	−0.44	−0.0455	−1.40 (0.200)

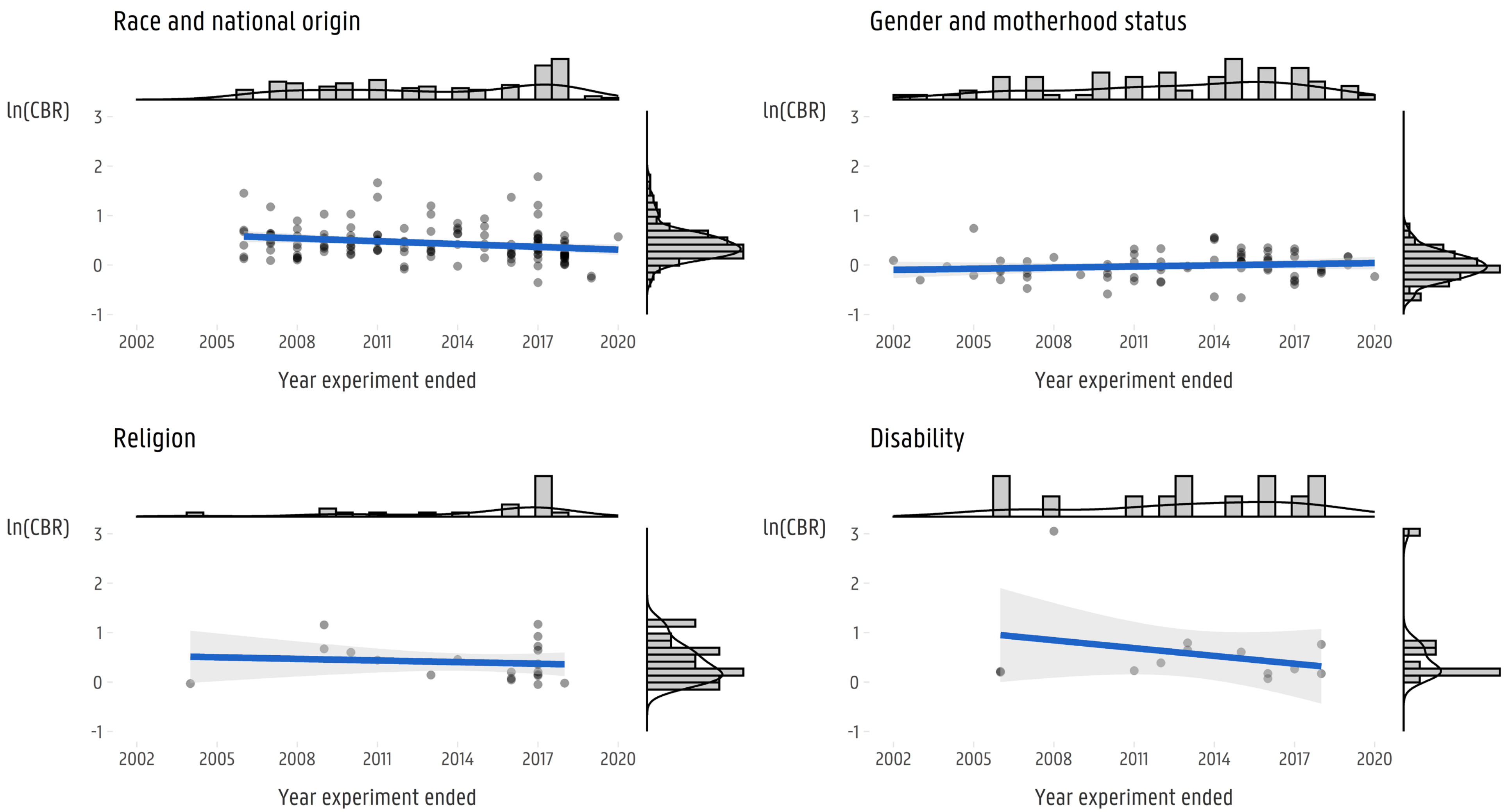
Notes. Abbreviations and notations used: *k* (number of correspondence experiments), $\hat{r}_{Pearson}$ (Pearson correlation coefficient), and LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations). The correlation coefficient is calculated as the weighted correlation between the majority/minority response ratios of the individual correspondence experiments and the year these experiments ended. Similarly, the regression coefficient is derived from the weighted least squares (WLS) model with said response ratios as the dependent variable and the year the experiments ended as the independent variable. We used the majority/minority response ratios because this allows us to interpret a negative (positive) correlation in terms of a decrease (increase) in hiring discrimination. Weights were derived from the meta-analytic random-effects model (see section 2.4.1). Following Harrer et al. (2021), statistics are only reported for discrimination grounds and treatment groups or regions nested within these grounds for which $k \geq 10$. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A11. Detected outliers by type of analysis

Discrimination ground or treatment group	Outliers
A. Main analyses	
Race and national origin	Agan & Starr (2016); Ahmad (2020); Andriessen et al. (2012); Arai et al. (2016); Arceo-Gomez & Campos-Vazques (2014); Asali et al. (2018); Attström (2007); Berson (2012); Bessudnov & Shcherbak (2020); Booth et al. (2012); Boyd-Swan & Herbst (2019); Bursell (2014); Busetta et al. (2018; 2020); Button & Walker (2020); Carlsson (2010); Cediey & Foroni (2008); Chowdhury et al. (2020); Decker et al. (2015); Derous et al. (2012); Duguet et al. (2010); Flake (2019); Gerhards et al. (2020); Jaeger et al. (2020); Kaas & Manger (2012); Koopmans et al. (2019); Lambert & Akinlade (2020); Lee & Khalid (2016); Nunley et al. (2015); Thijssen et al. (2021a; 2021b); Van den Berg et al. (2020); Veit & Thijssen (2021); Vernby & Dancygier (2019); Zschirnt (2020)
Gender and motherhood status	Ahmad (2020); Albert et al. (2011); Arceo-Gomez & Campos-Vazques (2014); Benhabib & Adair (2017); Booth & Leigh (2010); Busetta et al. (2018); Busetta et al. (2020); Capéau et al. (2012); Drydakis (2010a); Drydakis et al. (2017; 2018); Galarza & Yamada (2017); Gonzalez et al. (2019); Granberg et al. (2020); Mavlikeeva & Asanov (2020); Midtbøen (2016); Neumark et al. (2016); Oreopoulos & Dechief (2012); Oreopoulos (2011); Valfort (2020); Zhou et al. (2013)
Age	Baert (2014); Drydakis et al. (2017)
Religion	Drydakis (2010a)
Disability	Abubaker & Bagley (2017); Bessudnov & Shcherbak (2020); Drydakis (2010b); Hou et al. (2020); Koopmans et al. (2019); Weichselbaumer (2020); Yemane (2020)
Sexual orientation	Drydakis (2009; 2011; 2014)
Physical appearance	Busetta et al. (2021)
B. Sub-group analyses	
Race and national origin	
African/African American/Black	Agan & Starr (2016); Jaeger et al. (2020); Koopmans et al. (2019); Nunley et al. (2015)
Arab/Maghrebi/Middle Eastern	Ahmad (2020); Andriessen et al. (2012); Berson (2012); Duguet et al. (2010)
Eastern Asian/South-Eastern Asian	Lee & Khalid (2016)
Western Asian	Asali et al. (2018); Maurer-Fazio (2013)
Hispanic/Latin American/Caribbean	Flake (2019)
Mixed/Multiple	Lambert & Akinlade (2020)
Gender and motherhood status	
Female gender	Ahmad (2020); Albert et al. (2011); Arceo-Gomez & Campos-Vazques (2014); Booth & Leigh (2010); Busetta et al. (2018; 2020); Capéau et al. (2012); Drydakis (2010a); Drydakis et al. (2017; 2018); Galarza & Yamada (2017); Gonzalez et al. (2019); Horváth (2020); Mavlikeeva & Asanov (2020); Midtbøen (2016); Neumark et al. (2016); Valfort (2020); Zhou et al. (2013)
Age	
Old Age	Baert (2014); Drydakis et al. (2017)
Religion	
Muslim	Weichselbaumer (2020)
Disability	
Physical disability	Drydakis (2010a)
Sexual orientation	
LGB+ organisation affiliation	Drydakis (2009; 2014)

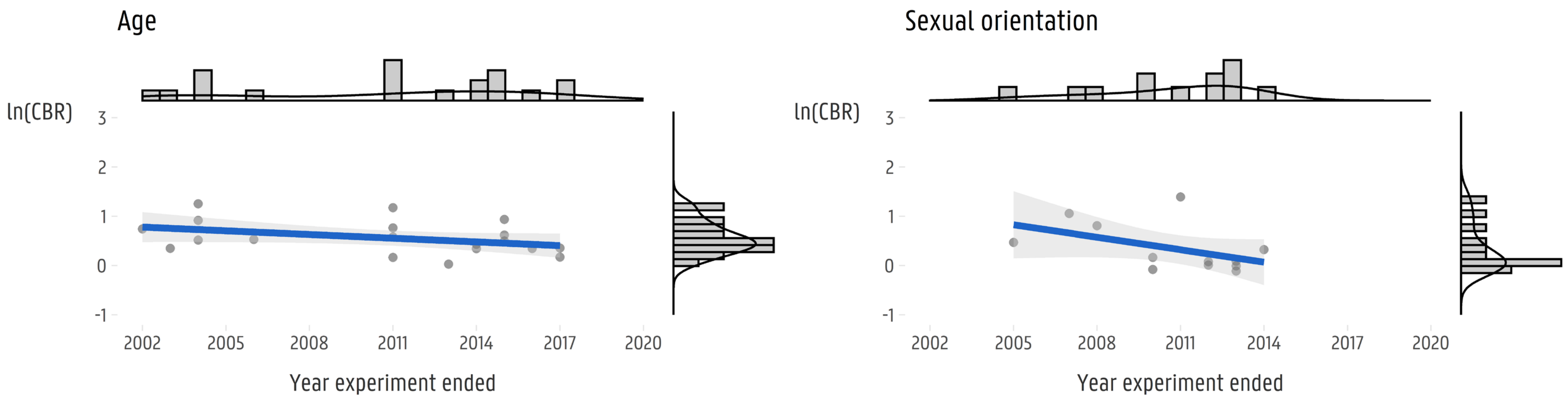
Notes. Abbreviations: LGB+ (lesbian, gay, and bisexual, amongst other sexual orientations). Following Harrer et al. (2021), outliers are defined as discrimination ratios for which the upper (lower) bound of the 95% confidence interval is lower (higher) than the lower (upper) bound of the confidence interval of the pooled discrimination ratio.

Figure A1. Correlation between the year in which the experiment ended and the log-adjusted majority/minority call-back ratios (regarding race and national origin, gender and motherhood, religion, and disability)



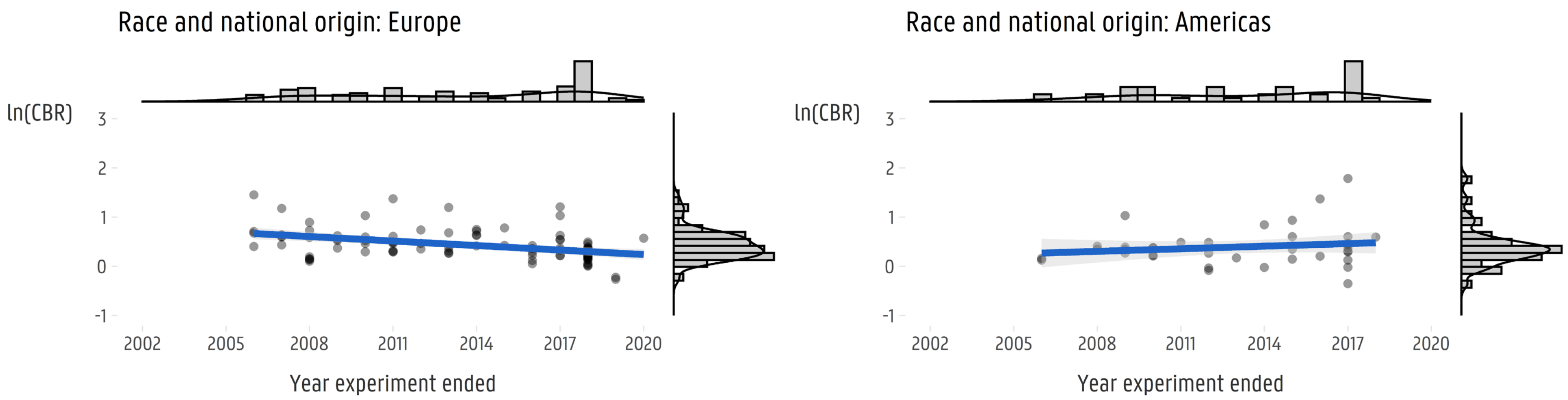
Notes. Marginal distributions consist of the histogram of the respective variables overlaid by the smoothed density of the distribution of these variables. Semi-transparent dots in the plots represent the natural log of the non-weighted majority/minority response ratios of the individual correspondence experiments. The dark lines and grey-marked areas represent the regression lines and their 95% confidence intervals based on the ordinary least squares model, respectively. Abbreviations used: ln(CBR) [log-adjusted majority/minority call-back ratios].

Figure A2. Correlation between the year in which the experiment ended and the log-adjusted majority/minority call-back ratios (regarding age and sexual orientation)



Notes. Marginal distributions consist of the histogram of the respective variables overlaid by the smoothed density of the distribution of these variables. Semi-transparent dots in the plots represent the natural log of the non-weighted majority/minority response ratios of the individual correspondence experiments. The dark lines and grey-marked areas represent the regression lines and their 95% confidence intervals based on the ordinary least squares model, respectively. Abbreviations used: ln(CBR) [log-adjusted majority/minority call-back ratios].

Figure A3. Correlation between the year in which the experiment ended and the log-adjusted majority/minority call-back ratios (regarding race and national origin distinguishing between European and American correspondence experiments).



Notes. Marginal distributions consist of the histogram of the respective variables overlaid by the smoothed density of the distribution of these variables. Semi-transparent dots in the plots represent the natural log of the non-weighted majority/minority response ratios of the individual correspondence experiments. The dark lines and grey-marked areas represent the regression lines and their 95% confidence intervals based on the ordinary least squares model, respectively. Abbreviations used: ln(CBR) [log-adjusted majority/minority call-back ratios].

Register of correspondence experiments

Table R1. Register of correspondence experiments on hiring discrimination published between 2005 and 2020

Treatment group	Control group	Country of analysis	Study	Effect
A. Race and national origin				
A.1. African American name	Anglo-Saxon name	United States of America	Jacquement & Yannelis (2012)	–
			Leasure & Andersen (2020)	0
			Vuolo et al. (2017)	–
A.2. African American or Somali American name	Anglo-Saxon name	United States of America	Gorsuch & Rho (2018)	–
A.3. African American origin	Anglo-Saxon origin	United States of America	Uggen et al. (2014)	–
A.4. African or Arab name	Native name	Sweden	Bursell (2014)	–
A.5. African or Hispanic name	Anglo-Saxon name	United States of America	Darolia et al. (2016)	0
			Decker et al. (2015)	0
A.6. African, Asian, or German name	Native name	Ireland	McGinnity & Lunn (2011)	–
A.7. African, Caribbean, Indian, or Pakistani name	Native name	United Kingdom of Great Britain and Northern Ireland	Wood et al. (2009)	–
A.8. Albanian name	Native name	Greece	Drydakis & Vlassis (2010)	–
			Drydakis (2012)	–
A.9. Albanian, Chinese, German, Moroccan, or Romanian name	Native name	Italy	Busetta et al. (2018)	–
A.10. Albanian, Georgian, or Ukrainian origin	Native origin	Greece	Drydakis (2017)	–
A.11. Antillean, Moroccan, Surinamese, or Turkish name	Native name	Netherlands	Andriessen et al. (2012)	–
A.12. Arab name	Anglo-Saxon name	United States of America	Widner & Chicoine (2011)	–
	Native name	Netherlands	Blommaert et al. (2014)	–
			Derous et al. (2012)	–
		Sweden	Aldén et al. (2021)	–
			Arai et al. (2016)	–
A.13. Arab origin	Native origin	France	Manant et al. (2019)	–
A.14. Armenian, Azeri, Chechen, Georgian, Latvian, Lithuanian, Uzbek, Tajik, or Tatar name	Native name	Russian Federation	Bessudnov & Shcherbak (2020)	–

A.15. Asian- or black-sounding name	Anglo-Saxon name	United States of America	Kang et al. (2016)	–
A.16. Asian domestic-born or white domestic-born	Asian foreign-born or white foreign-born	United States of America	Lambert & Akinlade (2020)	–
A.17. Asian phenotype	White phenotype	Germany	Koopmans et al. (2019)	0
A.18. Asian, Roma, or Turkish name	Native name	Czechia	Bartoš et al. (2016)	–
		Germany	Bartoš et al. (2016)	–
A.19. Azeri or Armenian name	Native name	Georgia	Asali et al. (2018)	–
A.20. Black- or Hispanic-sounding name	Anglo-Saxon name	United States of America	Boyd-Swan & Herbst (2019)	–
			Kleykamp (2009)	0
A.21. Black-, Hispanic-, or Asian-sounding name	Anglo-Saxon name	United States of America	Yemane (2020)	–
A.22. Black-sounding name	Anglo-Saxon name	United States of America	Agan & Starr (2016)	–
			Flake (2019)	–
			Gaddis (2015)	–
			Jaeger et al. (2020)	–
			Mobasseri (2019)	–
			Nunley et al. (2015)	–
A.23. Black phenotype or Turkish name	White phenotype or native name	Germany	Koopmans et al. (2019)	–
A.24. Chinese name	Anglo-Saxon name	Australia	Chowdhury et al. (2020)	–
A.25. Chinese, Greek, Indian, or Pakistani name	Anglo-Saxon name	Canada	Oreopoulos & Dechief (2012)	–
			Oreopoulos (2011)	–
A.26. Chinese, Indigenous, or Middle Eastern name	Native name	Australia	Booth et al. (2012)	–
A.27. Chinese, Nigerian, Serbian, or Turkish name and appearance	Native name and appearance	Austria	Weichselbaumer (2019)	–
A.28. Congolese, Italian, or Turkish name	Native name	Belgium	Capéau et al. (2012)	0
A.29. Cuban, Dominican, Ecuadorian, Mexican, Puerto Rican, or Salvadorian name	Anglo-Saxon or native name	Spain	Yemane & Fernández-Reino (2021)	0
		United States of America	Yemane & Fernández-Reino (2021)	–
A.30. Dark phenotype	Fair phenotype	Brazil	Dias (2020)	0
		Italy	Busetta et al. (2020)	–
		Pakistan	Saeed et al. (2019)	–
A.31. Eastern (European)-sounding name	Native name	Netherlands	Van den Berg et al. (2020)	–
			Thijssen et al. (2021a)	–
A.32. English, Iraqi, Russian, or Somali name	Native name	Finland	Ahmad (2020)	–
A.33. Foreign-born; Middle Eastern or African origin	Domestic-born; European origin	Germany	Veit & Thijssen (2021)	–

		Netherlands	Veit & Thijssen (2021)	–
		Norway	Veit & Thijssen (2021)	–
		Spain	Veit & Thijssen (2021)	0
		United Kingdom of Great Britain and Northern Ireland	Veit & Thijssen (2021)	–
A.34. French or German name	Native name	Switzerland	Zschirnt & Fibi (2019)	–
A.35. French or Turkish name	Native name	Germany	Gerhards et al. (2020)	0
A.36. German name	Native name	Switzerland	Zschirnt (2020)	0
A.37. German or Ukrainian name	Native name	Russian Federation	Bessudnov & Shcherbak (2020)	0
A.38. Ghanaian, Moroccan, Slovakian, or Turkish name	Native name	Belgium	Baert et al. (2017)	–
A.39. Hispanic-sounding name	Anglo-Saxon name	United States of America	Flake (2019)	+
			Mobasseri (2019)	0
	Native name	Netherlands	Thijssen et al. (2021a)	–
A.40. Indigenous-sounding name	Anglo-Saxon name	Peru	Galarza & Yamada (2017)	–
A.41. Indigenous origin	Native origin	United States of America	Button & Walker (2020)	0
A.42. Iraqi or Somali name	Native name	Sweden	Vernby & Dancygier (2019)	–
A.43. Italian name	Native name	Australia	Booth et al. (2012)	–
A.44. Kosovar name	Native name	Switzerland	Zschirnt & Fibi (2019)	–
			Zschirnt (2020)	–
A.45. Kurdish name	Native name	Turkey	Balkan & Cilasun (2018)	–
A.46. Malaysian name	Chinese name	Malaysia	Lee & Khalid (2016)	–
A.47. Mestizo phenotype	Fair phenotype	Mexico	Arceo-Gomez & Campos-Vazques (2014)	0
A.48. Middle Eastern name	Native name	Denmark	Dahl & Krog (2018)	–
			Guul et al. (2019)	–
			Villadsen & Wulff (2018)	–
		Netherlands	Thijssen et al. (2021)	–
		Sweden	Agerström et al. (2012)	–
			Attström (2007)	–
			Carlsson & Rooth (2007)	–
			Carlsson & Rooth (2012)	–
			Carlsson (2010)	–
A.49. Mongolian, Tibetan, or Uyghur name	Native name	China	Maurer-Fazio (2013)	–

A.50. Moroccan name	Native name	Belgium	Capéau et al. (2012)	–
		France	Berson (2012)	–
			Duguet et al. (2010)	–
			Pierné (2013)	–
		Netherlands	Ramos et al. (2021)	–
		Spain	Ramos et al. (2021)	–
A.51. Muslim Pakistani name	Native name	Norway	Birkelund et al. (2017)	–
A.52. North African name	Native name	France	Cahuc et al. (2019)	–
			Cediey & Foroni (2008)	–
			Edo et al. (2019)	–
A.53. Pakistani name	Native name	Norway	Birkelund et al. (2018)	–
			Larsen & Di Stasio (2021)	–
			Midtbøen (2016)	–
		United Kingdom of Great Britain and Northern Ireland	Larsen & Di Stasio (2021)	–
A.54. Polish name	Native name	Sweden	Vernby & Dancygier (2019)	0
A.55. South Asian name	Native name	Netherlands	Thijssen et al. (2021a)	0
A.56. Southeast or East Asian name	Native name	Netherlands	Thijssen et al. (2021a)	–
A.57. Southern European name	Native name	Netherlands	Thijssen et al. (2021a)	–
A.58. Turkish name	Native name	Belgium	Baert & Vujić (2016)	–
			Baert et al. (2015)	–
		Germany	Kaas & Manger (2012)	–
			Thijssen et al. (2021b)	–
			Weichselbaumer (2020)	–
		Netherlands	Thijssen et al. (2021b)	–
		Switzerland	Zschirnt & Fibi (2019)	0
			Zschirnt (2020)	0
A.59. Western (American or Western European) name	Native name	Netherlands	Thijssen et al. (2021a)	–
B. Gender and motherhood status				
B.1. Mother	Childless woman	Austria	Becker et al. (2019)	0
		Belgium	Baert (2014)	0
		Germany	Becker et al. (2019)	0
			Hipp (2020)	–

B.1. Mother	Childless woman	Spain	Albert et al. (2011)	0
			Gonzalez et al. (2019)	–
		Sweden	Bygren et al. (2017)	0
		Switzerland	Becker et al. (2019)	0
		United States of America	Correll et al. (2007)	–
B.2. Pregnant	No pregnancy	Belgium	Capéau et al. (2012)	0
B.3. Female gender	Male gender	Algeria	Benhabib & Adair (2017)	+
		Australia	Booth & Leigh (2010)	+
			Booth et al. (2012)	0
		Belgium	Baert et al. (2016a)	0
			Baert et al. (2016b)	0
			Baert et al. (2016c)	0
			Capéau et al. (2012)	–
		Canada	Oreopoulos & Dechief (2012)	+
			Oreopoulos (2011)	+
		China	Deng et al. (2020)	+
			Horvath (2020)	0
			Maurer-Fazio & Lei (2015)	+
			Wu (2017)	0
			Zhou et al. (2013)	+
		Cyprus	Drydakis (2014)	0
		Denmark	Dahl & Krog (2018)	0
		Finland	Ahmad (2020)	+
		France	Berson (2012)	+
			Duguet et al. (2017)	–
			Edo et al. (2019)	+
			Petit (2007)	0
			Valfort (2020)	+
		Georgia	Asali et al. (2018)	0
		Germany	Hipp (2020)	0
		Greece	Drydakis (2010a)	–
		Greece	Drydakis (2017)	–
		India	Banerjee et al. (2009)	0

B.3. Female gender	Male gender	Israel	Ruffle & Shtudiner (2015)	0
		Italy	Busetta et al. (2018)	−
			Busetta et al. (2020)	0
			Busetta et al. (2021)	0
			Patacchini et al. (2015)	0
		Mexico	Arceo-Gomez & Campos-Vazques (2014)	+
			Campos-Vazquez & Gonzalez (2020)	+
		Netherlands	Ramos et al. (2021)	0
		Norway	Midtbøen (2016)	+
		Pakistan	Saeed et al. (2019)	0
		Peru	Galarza & Yamada (2017)	−
		Russian Federation	Mavlikeeva & Asanov (2020)	+
		Spain	Albert et al. (2011)	+
			Gonzalez et al. (2019)	−
			Ramos et al. (2021)	0
		Sweden	Ahmed et al. (2013)	0
			Aldén et al. (2021)	0
			Arai et al. (2016)	0
			Attström (2007)	0
			Bailey et al. (2013)	0
			Bygren et al. (2017)	0
			Carlsson & Eriksson (2019)	0
			Carlsson (2011)	0
			Carlsson et al. (2014)	0
			Erlandsson (2019)	−
			Rooth (2009)	0
		Switzerland	Zschirnt (2020)	0
		Turkey	Balkan & Cilasun (2018)	0
		United Kingdom of Great Britain and Northern Ireland	Drydakis (2015)	0
			Drydakis et al. (2017)	−
			Jackson (2009)	+
			Riach & Rich (2006)	+

B.3. Female gender	Male gender	United States of America	Correll et al. (2007)	0
			Neumark et al. (2016)	–
			Neumark et al. (2019)	+
			Rivera & Tilcsik (2016)	0
			Yemane & Fernández-Reino (2021)	+
B.4. Female gender with feminine personality traits	Female gender with masculine personality traits	United Kingdom of Great Britain and Northern Ireland	Drydakis et al. (2018)	–
B.5. Transgender	Cisgender	Germany	Gerhards et al. (2020)	0
		Sweden	Granberg et al. (2020)	–
C. Religion				
C.1. Buddhist, Christian or Hindu	No religious affiliation	United States of America	Yemane (2020)	0
C.2. Disclosed Muslims	Muslims by default	Germany	Di Stasio et al. (2021)	0
		Netherlands	Di Stasio et al. (2021)	–
		Norway	Di Stasio et al. (2021)	–
		Spain	Di Stasio et al. (2021)	0
		United Kingdom of Great Britain and Northern Ireland	Di Stasio et al. (2021)	–
C.3. Evangelical, Jehovah's Witness, or Pentecostal	Majority religion	Greece	Drydakis (2010b)	–
C.4. Expressing a religious identity	Expressing no religious identity	United States of America	Wright et al. (2014)	–
C.5. Hui or Uyghur Muslim	Han Chinese	China	Hou et al. (2020)	–
C.6. Jewish	No religious affiliation	Russian Federation	Bessudnov & Shcherbak (2020)	0
C.7. Muslim	Christian	Germany	Koopmans et al. (2019)	–
	Majority religion	France	Adida et al. (2010)	–
			Pierné (2013)	–
			Valfort (2020)	–
		India	Banerjee et al. (2009)	0
		Netherlands	Abubaker & Bagley (2017)	–
		United Kingdom of Great Britain and Northern Ireland	Abubaker & Bagley (2017)	–
		United States of America	Acquisti & Fong (2020)	0
	No religious affiliation	United States of America	Yemane (2020)	–
C.8. No religious affiliation	Christian	Germany	Koopmans et al. (2019)	0
C.9. Wearing a headscarf	Not wearing a headscarf	Germany	Weichselbaumer (2020)	–

D. Disability				
D.1. Asperger's Syndrome or spinal cord injury	No disability	United States of America	Ameri et al. (2018)	0
D.2. Autism, blindness, or deafness	No disability	Belgium	Baert (2016)	–
D.3. Obese	Not obese	Italy	Busetta et al. (2020)	–
		Mexico	Campos-Vazquez & Gonzalez (2020)	–
		Sweden	Agerström & Rooth (2011)	–
		Sweden	Rooth (2009)	–
D.4. Former depression	No former depression	Belgium	Baert et al. (2016b)	–
D.5. History of mental illness	Physical injury	United States of America	Hipes et al. (2016)	–
D.6. HIV infection	No HIV infection	Greece	Drydakis (2010a)	–
D.7. Unspecified physical disability	No physical disability	Belgium	Capéau et al. (2012)	0
D.8. Wheelchair user	No physical disability	Canada	Bellemare et al. (2020)	–
		United Kingdom of Great Britain and Northern Ireland	Stone & Wright (2013)	–
E. Age				
E.1. Age 22–23 or age 60–61	Age 33–34 or age 42–43	United States of America	Farber et al. (2019)	–
E.2. Age 23	Age 35	Belgium	Capéau et al. (2012)	0
E.3. Age 36–70	Age 35–69	Sweden	Carlsson & Eriksson (2019)	–
E.4. Age 37	Age 25	Belgium	Baert (2014)	0
		France	Petit (2007)	–
E.5. Age 38	Age 24 or age 28	Spain	Albert et al. (2011)	–
E.6. Age 46	Age 31	Sweden	Ahmed et al. (2012)	–
E.7. Age 47	Age 27	France	Riach (2015)	–
		Germany	Riach (2015)	–
		Spain	Riach (2015)	–
		United Kingdom of Great Britain and Northern Ireland	Riach (2015)	–
E.8. Age 47 or age 53	Age 35	Belgium	Capéau et al. (2012)	–
E.9. Age 49–51 or age 64–66	Age 29–31	United States of America	Neumark et al. (2016)	–
E.10. Age 50	Age 28	United Kingdom of Great Britain and Northern Ireland	Drydakis et al. (2017)	–
E.11. Age 50 or age 44	Age 44 or age 38	Belgium	Baert et al. (2016c)	–
E.12. Age 50 or age 51	Age 24 or age 25	United Kingdom of Great Britain and Northern Ireland	Tinsley (2012)	–

E.13. Age 50, age 55, or age 62	Age 35 or age 45	United States of America	Lahey (2008)	–
E.14. Age 56	Age 29	France	Challe et al. (2015)	–
E.15. Age 57 or age 58	Age 37 or age 42	United States of America	Farber et al. (2016)	–
E.16. Age 64–66	Age 29–31	United States of America	Neumark et al. (2019)	–
F. Military service or affiliation				
F.1. Current membership in the Reserves	Previous membership in the reserves	United States of America	Figinski (2017)	–
F.2. Military service	No military service	Belgium	Baert & Balcaen (2013)	0
		United States of America	Figinski (2019)	+
			Kleykamp (2009)	0
G. Wealth				
G.1. Lower-class background	Higher-class background	Jamaica	Spencer et al. (2019)	0
		United States of America	Rivera & Tilcsik (2016)	–
			Thomas (2018)	0
G.2. Non-upper-caste	Upper-caste	India	Banerjee et al. (2009)	0
			Siddique (2011)	0
G.3. Residence in neighbourhood with bad or poor reputation	Residence in neighbourhood with good or bland reputation	Sweden	Carlsson et al. (2018)	0
		France	Bunel et al. (2016)	–
		United Kingdom of Great Britain and Northern Ireland	Tunstall et al. (2014)	0
H. Genetic information				
No related correspondence experiments found.				
I. Citizenship status				
No related correspondence experiments found.				
J. Marital status				
J.1. Married	Not married	China	Horvath (2020)	0
		Germany	Weichselbaumer (2015)	–
		Mexico	Arceo-Gomez & Campos-Vazques (2014)	0
J.2. Married and childless	Single and childless	Austria	Becker et al. (2019)	0
		Germany	Becker et al. (2019)	0
		Switzerland	Becker et al. (2019)	0
K. Sexual orientation				
K.1. LGB+ organisation affiliation	No LGB+ organisation affiliation	Cyprus	Drydakis (2014)	–
		Germany	Weichselbaumer (2015)	–

		Greece	Drydakis (2009)	−
			Drydakis (2011)	−
		Italy	Patacchini et al. (2015)	0
		Sweden	Ahmed et al. (2013)	−
			Bailey et al. (2013)	0
		United Kingdom of Great Britain and Northern Ireland	Drydakis (2015)	−
		United States of America	Mishel (2016)	−
			Tilcsik (2011)	−
K.2. Same-sex marriage partner	Opposite sex marriage partner	Belgium	Baert (2014)	0
K.3. Same-sex orientation	Opposite sex orientation	United States of America	Acquisti & Fong (2020)	0
L. Political orientation				
L.1. Orientation of mentioned youth political organisation	Orientation of mentioned youth political organisation	Belgium	Baert et al. (2014)	0
M. Union membership				
M.1. Mention of youth union membership	No mention of youth union membership	Belgium	Baert & Omey (2015)	−
N. Physical appearance				
N.1. Facial disfigurement in resume picture	No facial disfigurement in resume picture	United Kingdom of Great Britain and Northern Ireland	Stone & Wright (2013)	−
N.2. Visible tattoo	No (visible) tattoo	Germany	Jibuti (2018)	−
N.3. Lower physical attractiveness of resume picture	Higher physical attractiveness of resume picture	Argentina	Lopez Bóo et al. (2013)	−
		Belgium	Baert (2018)	−
		China	Deng et al. (2020)	−
			Maurer-Fazio & Lei (2015)	−
			Wu (2017)	−
		Israel	Ruffle & Shtudiner (2015)	−
		Italy	Busetta et al. (2021)	−
			Patacchini et al. (2015)	0
		Peru	Galarza & Yamada (2017)	−
		Philippines	Beam et al. (2020)	−

Notes. Values in the 'Effect' column illustrate the effect of the treatment group on the call-back variable in the respective correspondence experiments. The +, 0, and − symbol indicate an overall significantly positive, neutral, and negative effect, respectively.

List of studies included in the register of correspondence experiments

[1] Abubaker, M., & Bagley, C. A. (2017). Methodology of correspondence testing for employment discrimination involving ethnic minority applications: Dutch and English case studies of Muslim applicants for employment. *Social Sciences, 6*(4), 112–121. https://doi.org/10.3390/socsci6040112

[2] Acquisti, A., & Fong, C. (2020). An experiment in hiring discrimination via online social networks. *Management Science, 66*(3), 1005–1024. https://doi.org/10.1287/mnsc.2018.3269

[3] Adida, C. L., Laitin, D. D., & Valfort, M.-A. (2010). Identifying barriers to Muslim integration in France. *Proceedings of the National Academy of Sciences, 107*(52), 22384–22390. https://doi.org/10.1073/pnas.1015550107

[4] Agan, A. Y., & Starr, S. B. (2016). *Ban the box, criminal records, and statistical discrimination: A field experiment* (SSRN Electronic Journal Working Paper No. 16–012). University of Chicago, Becker Friedman Institute for Economics. https://doi.org/10.2139/ssrn.2795795

[5] Agerström, J., & Rooth, D.-O. (2011). The role of automatic obesity stereotypes in real hiring discrimination. *Journal of Applied Psychology, 96*(4), 790–805. https://doi.org/10.1037/a0021594

[6] Agerström, J., Björklund, F., Carlsson, R., & Rooth, D.-O. (2012). Warm and competent Hassan = cold and incompetent Eric: A harsh equation of real-life hiring discrimination. *Basic and Applied Social Psychology, 34*(4), 359–366. https://doi.org/10.1080/01973533.2012.693438

[7] Ahmad, A. (2019). When the name matters: An experimental investigation of ethnic discrimination in the Finnish labor market. *Sociological Inquiry, 90*(3), 468–496. https://doi.org/10.1111/soin.12276

[8] Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2012). Does age matter for employability? A field experiment on ageism in the Swedish labour market. *Applied Economics Letters, 19*(4), 403–406. https://doi.org/10.1080/13504851.2011.581199

[9] Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2013). Are gay men and lesbians discriminated against in the hiring process? *Southern Economic Journal, 79*(3), 565–585.

https://doi.org/10.4284/0038-4038-2011.317

[10] 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. https://doi.org/10.1080/09585192.2011.540160

[11] Aldén, L., Bastani, S., & Hammarstedt, M. (2021). Ethnic background and the value of self-employment experience: Evidence from a randomized field experiment. *Oxford Bulletin of Economics and Statistics, 0305-9049,* 1–24. https://doi.org/10.1111/obes.12454

[12] Ameri, M., Schur, L., Adya, M., Bentley, F. S., McKay, P., & Kruse, D. (2017). The disability employment puzzle: A field experiment on employer hiring behavior. *ILR Review, 71*(2), 329–364. https://doi.org/10.1177/0019793917717474

[13] Andriessen, I., Nievers, E., Dagevos, J., & Faulk, L. (2012). Ethnic discrimination in the Dutch labor market. *Work and Occupations, 39*(3), 237–269. https://doi.org/10.1177/0730888412444783

[14] Arai, M., Bursell, M., & Nekby, L. (2016). The reverse gender gap in ethnic discrimination: Employer stereotypes of men and women with Arabic names. *International Migration Review, 50*(2), 385–412. https://doi.org/10.1111/imre.12170

[15] Arceo-Gomez, E. O., & Campos-Vazquez, R. M. (2014). Race and marriage in the labor market: A discrimination correspondence study in a developing country. *American Economic Review, 104*(5), 376–380. https://doi.org/10.1257/aer.104.5.376

[16] 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. https://doi.org/10.1016/j.jce.2018.09.001

[17] Attström, K., & Attström, K. (2007*). Discrimination against native Swedes of immigrant origin in access to employment: A research study on Stockholm, Malmö and Gothenburg, utilizing the ILO discrimination practice testing approach* (International Migration Papers 86E). International Labour Organization. https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---migrant/documents/publication/wcms_201428.pdf

[18] Baert, S. (2014a). Career lesbians. Getting hired for not having kids? *Industrial Relations Journal, 45*(6), 543–561. https://doi.org/10.1111/irj.12078

[19] Baert, S. (2014b). Wage subsidies and hiring chances for the disabled: Some causal evidence. *The European Journal of Health Economics, 17*(1), 71–86.

https://doi.org/10.1007/s10198-014-0656-7

[20] Baert, S. (2017). Facebook profile picture appearance affects recruiters' first hiring decisions. *New Media & Society, 20*(3), 1220–1239. https://doi.org/10.1177/1461444816687294

[21] Baert, S., & Balcaen, P. (2013). The impact of military work experience on later hiring chances in the civilian labour market. Evidence from a Field Experiment. *Economics, 7*(1). https://doi.org/10.5018/economics-ejournal.ja.2013-37

[22] Baert, S., Jong-A-Pin, R., De Freyne, L., & Parmentier, S. (2014, April 24–26). *Political ideology and labour market discrimination* [Conference presentation]. Spring Meeting of Young Economists 2014, Vienna, Austria. https://users.ugent.be/~sbaert/research_results.htm

[23] Baert, S., & Omey, E. (2015). Hiring discrimination against pro-union applicants: The role of union density and firm size. *De Economist, 163*(3), 263–280. https://doi.org/10.1007/s10645-015-9252-1

[24] Baert, S., & Vujić, S. (2016). Immigrant volunteering: A way out of labour market discrimination? *Economics Letters, 146*, 95–98. https://doi.org/10.1016/j.econlet.2016.07.035

[25] Baert, S., Albanese, A., du Gardein, S., Ovaere, J., & Stappers, J. (2017). Does work experience mitigate discrimination? *Economics Letters, 155*, 35–38. https://doi.org/10.1016/j.econlet.2017.03.011

[26] Baert, S., Cockx, B., Gheyle, N., & Vandamme, C. (2015). Is there less discrimination in occupations where recruitment is difficult? *ILR Review, 68*(3), 467–500. https://doi.org/10.1177/0019793915570873

[27] Baert, S., De Pauw, A.-S., & Deschacht, N. (2016). Do employer preferences contribute to sticky floors? *ILR Review, 69*(3), 714–736. https://doi.org/10.1177/0019793915625213

[28] Baert, S., De Visschere, S., Schoors, K., Vandenberghe, D., & Omey, E. (2016). First depressed, then discriminated against? *Social Science & Medicine, 170*, 247–254. https://doi.org/10.1016/j.socscimed.2016.06.033

[29] Baert, S., Norga, J., Thuy, Y., & Van Hecke, M. (2016). Getting grey hairs in the labour market. An alternative experiment on age discrimination. *Journal of Economic Psychology, 57*, 86–101. https://doi.org/10.1016/j.joep.2016.10.002

[30] Bailey, J., Wallace, M., & Wright, B. (2013). Are gay men and lesbians discriminated against when applying for jobs? A four-city, internet-based field experiment. *Journal of Homosexuality, 60*(6), 873–894. https://doi.org/10.1080/00918369.2013.774860

[31] Balkan, B., & Cilasun, S. M. (2018). *Ethnic discrimination in the Turkish labor market: Evidence from survey and field data* (Economic Research Forum Working Working Paper No. 1197). The Economic Research Forum. https://erf.org.eg/publications/ethnic-discrimination-in-the-turkish-labor-market-evidence-from-survey-and-field-data

[32] Banerjee, A., Bertrand, M., Datta, S., & Mullainathan, S. (2009). Labor market discrimination in Delhi: Evidence from a field experiment. *Journal of Comparative Economics, 37*(1), 14–27. https://doi.org/10.1016/j.jce.2008.09.002

[33] Bartoš, V., Bauer, M., Chytilová, J., & Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review, 106*(6), 1437–1475. https://doi.org/10.1257/aer.20140571

[34] Beam, E. A., Hyman, J., & Theoharides, C. (2020). The relative returns to education, experience, and attractiveness for young workers. *Economic Development and Cultural Change, 68*(2), 391–428. https://doi.org/10.1086/701232

[35] Becker, S. O., Fernandes, A., & Weichselbaumer, D. (2019). Discrimination in hiring based on potential and realized fertility: Evidence from a large-scale field experiment. *Labour Economics, 59*, 139–152. https://doi.org/10.1016/j.labeco.2019.04.009

[36] Bellemare, C., Goussé, M., Lacroix, G., & Marchand, S. (2020). *Video resumes and job search outcomes: Evidence from a field experiment* (IZA Discussion Papers No. 13656). IZA Institute of Labor Economics. https://ideas.repec.org/p/iza/izadps/dp13656.html

[37] Benhabib, L., & Adair, P. (2017, June 29–July 1). *Hiring discrimination on the Algerian labour market: An assessment with testing* [Conference presentation]. Africa Meeting of the Econometric Society 2017, Algiers, Algeria. https://hal.archives-ouvertes.fr/hal-01683421/

[38] Berson, C. (2012). *Does competition induce hiring equity?* (Documents de travail du Centre d'Economie de la Sorbonne No. 12019). Université Panthéon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne. https://ideas.repec.org/p/mse/cesdoc/12019.html

[39] Bessudnov, A., & Shcherbak, A. (2019). Ethnic discrimination in multi-ethnic societies: Evidence from russia. *European Sociological Review, 36*(1), 104–120. https://doi.org/10.1093/esr/jcz045

[40] Birkelund, G. E., Chan, T. W., Ugreninov, E., Midtbøen, A. H., & Rogstad, J. (2018). Do terrorist attacks affect ethnic discrimination in the labour market? Evidence from two randomized field experiments. *The British Journal of Sociology, 70*(1), 241–260. https://doi.org/10.1111/1468-4446.12344

[41] Birkelund, G. E., Heggebø, K., & Rogstad, J. (2016). Additive or multiplicative disadvantage? The scarring effects of unemployment for ethnic minorities. *European Sociological Review, 33*(1), 17-29. https://doi.org/10.1093/esr/jcw030

[42] Blommaert, L., Coenders, M., & van Tubergen, F. (2013). Discrimination of Arabic-named applicants in the Netherlands: An internet-based field experiment examining different phases in online recruitment procedures. *Social Forces, 92*(3), 957–982. https://doi.org/10.1093/sf/sot124

[43] Booth, A. L., Leigh, A., & Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics, 74*(4), 547–573. https://doi.org/10.1111/j.1468-0084.2011.00664.x

[44] Booth, A., & Leigh, A. (2010). Do employers discriminate by gender? A field experiment in female-dominated occupations. *Economics Letters, 107*(2), 236–238. https://doi.org/10.1016/j.econlet.2010.01.034

[45] Boyd-Swan, C., & Herbst, C. M. (2019). Racial and ethnic discrimination in the labor market for child care teachers. *Educational Researcher, 48*(7), 394–406. https://doi.org/10.3102/0013189x19867941

[46] Bunel, M., L'Horty, Y., & Petit, P. (2015). Discrimination based on place of residence and access to employment. *Urban Studies, 53*(2), 267–286. https://doi.org/10.1177/0042098014563470

[47] Bursell, M. (2014). The multiple burdens of foreign-named men: Evidence from a field experiment on gendered ethnic hiring discrimination in Sweden. *European Sociological Review, 30*(3), 399–409. https://doi.org/10.1093/esr/jcu047

[48] Busetta, G., Campolo, M. G., & Panarello, D. (2018). Immigrants and Italian labor market: Statistical or taste-based discrimination? *Genus, 74*(1). https://doi.org/10.1186/s41118-018-0030-1

[49] Busetta, G., Campolo, M. G., & Panarello, D. (2020). Weight-based discrimination in the Italian labor market: An analysis of the interaction with gender and ethnicity. *The Journal of Economic Inequality, 18*(4), 617–637. https://doi.org/10.1007/s10888-020-09456-5

[50] Busetta, G., Fiorillo, F., & Palomba, G. (2020). The impact of attractiveness on job opportunities in Italy: A gender field experiment. *Economia Politica, 38*(1), 171–201. https://doi.org/10.1007/s40888-020-00194-5

[51] Button, P., & Walker, B. (2020). Employment discrimination against Indigenous Peoples in the United States: Evidence from a field experiment. *Labour Economics, 65*, 101851. https://doi.org/10.1016/j.labeco.2020.101851

[52] 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. https://doi.org/10.1093/esr/jcx051

[53] Cahuc, P., Carcillo, S. L., Minea, A., & Valfort, M. A. (2019). *When correspondence studies fail to detect hiring discrimination* (IZA Discussion Papers No. 12653). IZA Institute of Labor Economics. https://ideas.repec.org/p/iza/izadps/dp12653.html

[54] Campos-Vazquez, R. M., & Gonzalez, E. (2020). Obesity and hiring discrimination. *Economics & Human Biology, 37*, 100850. https://doi.org/10.1016/j.ehb.2020.100850

[55] Capéau, B., Eeman, L., Groenez, S., & Lamberts, M. (2012). *Two concepts of discrimination: inequality of opportunity versus unequal treatment of equals* (European Center for Advanced Research in Economics and Statistics [ECARES] Working Papers No. 2012–021). ECARES, Université Libre de Bruxelles. https://econpapers.repec.org/paper/ecawpaper/2013_2f123943.htm

[56] Carlsson, M. (2010). Experimental evidence of discrimination in the hiring of first- and second-generation immigrants. *Labour, 24*(3), 263–278. https://doi.org/10.1111/j.1467-9914.2010.00482.x

[57] Carlsson, M. (2011). Does hiring discrimination cause gender segregation in the Swedish labor market? *Feminist Economics, 17*(3), 71–102. https://doi.org/10.1080/13545701.2011.580700

[58] Carlsson, M., & Eriksson, S. (2019). Age discrimination in hiring decisions: Evidence from a field experiment in the labor market. *Labour Economics, 59*, 173–183. https://doi.org/10.1016/j.labeco.2019.03.002

[59] Carlsson, M., & Rooth, D.-O. (2007). Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics, 14*(4), 716–729. https://doi.org/10.1016/j.labeco.2007.05.001

[60] Carlsson, M., & Rooth, D.-O. (2012). Revealing taste-based discrimination in hiring: a correspondence testing experiment with geographic variation. *Applied Economics Letters, 19*(18), 1861–1864. https://doi.org/10.1080/13504851.2012.667537

[61] Carlsson, M., Reshid, A. A., & Rooth, D.-O. (2018). Neighborhood signaling effects, commuting time, and employment. *International Journal of Manpower, 39*(4), 534–549. https://doi.org/10.1108/ijm-09-2017-0234

[62] Carlsson, R., Agerström, J., Björklund, F., Carlsson, M., & Rooth, D.-O. (2014). Testing for backlash in hiring. *Journal of Personnel Psychology, 13*(4), 204–214. https://doi.org/10.1027/1866-5888/a000126

[63] Cediey, E., & Foroni, F. (2008). *Discrimination in access to employment on grounds of foreign origin in France: A national survey of discrimination based on the testing methodology of the International Labour Office* (International Migration Papers 85E). International Labour Organization. https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---migrant/documents/publication/wcms_201429.pdf

[64] Challe, L., Fremigacci, F., Langot, F., l'Horty, Y., Du Parquet, L., & Petit, P. (2015). *Access to employment with age and gender: Results of a controlled experiment* (HAL Working Papers No. 01292137). HAL, Paris School of Economics. https://hal.archives-ouvertes.fr/hal-01292137

[65] Chowdhury, S., Ooi, E., & Slonim, R. (2020). *Racial discrimination and white first name adoption: Evidence from a correspondence study in the Australian labour market* (IZA Discussion Papers No. 13208). IZA Institute of Labor Economics. https://ideas.repec.org/p/iza/izadps/dp13208.html

[66] Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology, 112*(5), 1297–1339. https://doi.org/10.1086/511799

[67] Dahl, M., & Krog, N. (2018). Experimental evidence of discrimination in the labour market: Intersections between ethnicity, gender, and socio-economic status. *European Sociological Review, 34*(4), 402–417. https://doi.org/10.1093/esr/jcy020

[68] Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2015). Race and gender effects on employer interest in job applicants: New evidence from a resume field experiment. *Applied Economics Letters, 23*(12), 853–856. https://doi.org/10.1080/13504851.2015.1114571

[69] Decker, S. H., Ortiz, N., Spohn, C., & Hedberg, E. (2015). Criminal stigma, race, and

ethnicity: The consequences of imprisonment for employment. *Journal of Criminal Justice, 43*(2), 108–121. https://doi.org/10.1016/j.jcrimjus.2015.02.002

[70] Deng, W., Li, D., & Zhou, D. (2019). Beauty and job accessibility: New evidence from a field experiment. *Journal of Population Economics, 33*(4), 1303–1341. https://doi.org/10.1007/s00148-019-00744-7

[71] Derous, E., Ryan, A. M., & Nguyen, H.-H. D. (2011). Multiple categorization in resume screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings. *Journal of Organizational Behavior, 33*(4), 544–570. https://doi.org/10.1002/job.769

[72] Di Stasio, V., Lancee, B., Veit, S., & Yemane, R. (2019). 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. https://doi.org/10.1080/1369183x.2019.1622826

[73] Dias, F. A. (2020). How skin color, class status, and gender intersect in the labor market: Evidence from a field experiment. *Research in Social Stratification and Mobility, 65*, 100477. https://doi.org/10.1016/j.rssm.2020.100477

[74] Drydakis, N. (2009). Sexual orientation discrimination in the labour market. *Labour Economics, 16*(4), 364–372. https://doi.org/10.1016/j.labeco.2008.12.003

[75] Drydakis, N. (2010a). Labour discrimination as a symptom of HIV: Experimental evaluation – the Greek case. *Journal of Industrial Relations, 52*(2), 201–217. https://doi.org/10.1177/0022185609359445

[76] Drydakis, N. (2010b). Religious affiliation and employment bias in the labor market. *Journal for the Scientific Study of Religion, 49*(3), 477–493. https://doi.org/10.1111/j.1468-5906.2010.01523.x

[77] Drydakis, N. (2011). Women's sexual orientation and labor market outcomes in Greece. *Feminist Economics, 17*(1), 89–117. https://doi.org/10.1080/13545701.2010.541858

[78] Drydakis, N. (2012). Estimating ethnic discrimination in the labour market using experimental data. *Southeast European and Black Sea Studies, 12*(2), 335–355. https://doi.org/10.1080/14683857.2012.686016

[79] Drydakis, N. (2014). Sexual orientation discrimination in the Cypriot labour market. Distastes or uncertainty? *International Journal of Manpower, 35*(5), 720–744.

https://doi.org/10.1108/ijm-02-2012-0026

[80] Drydakis, N. (2015). Sexual orientation discrimination in the United Kingdom's labour market: A field experiment. *Human Relations, 68*(11), 1769–1796. https://doi.org/10.1177/0018726715569855

[81] Drydakis, N. (2017). Measuring labour differences between natives, non-natives, and natives with an ethnic-minority background. *Economics Letters, 161*, 27–30. https://doi.org/10.1016/j.econlet.2017.08.031

[82] Drydakis, N., & Vlassis, M. (2010). Ethnic discrimination in the Greek labour market: Occupational access, insurance coverage and wage offers. *The Manchester School, 78*(3), 201–218. https://doi.org/10.1111/j.1467-9957.2009.02132.x

[83] Drydakis, N., MacDonald, P., Bozani, V., & Chiotis, V. (2017). *Inclusive recruitment? Hiring discrimination against older workers.* In A. Arenas, D. Di Marco, L. Munduate, & M. Euwema (Eds.), *Shaping inclusive workplaces through social dialogue. Industrial relations & conflict management* (Vol. 1, pp. 87–102). Springer, Cham. https://doi.org/10.1007/978-3-319-66393-7_6

[84] Drydakis, N., Sidiropoulou, K., Bozani, V., Selmanovic, S., & Patnaik, S. (2018). Masculine vs feminine personality traits and women's employment outcomes in Britain. *International Journal of Manpower, 39*(4), 621–630. https://doi.org/10.1108/ijm-09-2017-0255

[85] Duguet, E., Leandri, N., L'Horty, Y., & Petit, P. (2010). Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area. *Annals of Economics and Statistics, 99*, 187–215. https://doi.org/10.2307/41219165

[86] Duguet, E., Looc, D. P., & Petit, P. (2017). *Hiring discrimination against women: Distinguishing taste based discrimination from statistical discrimination* (SSRN Electronic Journal Working Paper No. 3083957). University of Chicago, Becker Friedman Institute for Economics. https://doi.org/10.2139/ssrn.3083957

[87] Edo, A., Jacquemet, N., & Yannelis, C. (2017). Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications. *Review of Economics of the Household, 17*(1), 349–376. https://doi.org/10.1007/s11150-017-9391-z

[88] Erlandsson, A. (2019). Do men favor men in recruitment? A field experiment in the Swedish labor market. *Work and Occupations, 46*(3), 239–264. https://doi.org/10.1177/0730888419849467

[89] Farber, H. S., Herbst, C. M., Silverman, D., & von Wachter, T. (2019). Whom do employers want? The role of recent employment and unemployment status and age. *Journal of Labor Economics, 37*(2), 323–349. https://doi.org/10.1086/700184

[90] Farber, H. S., Silverman, D., & von Wachter, T. (2016). Determinants of callbacks to job applications: An audit study. *American Economic Review, 106*(5), 314–318. https://doi.org/10.1257/aer.p20161010

[91] Figinski, T. F. (2016). The effect of potential activations on the employment of military reservists: Evidence from a field experiment. *ILR Review, 70*(4), 1037–1056. https://doi.org/10.1177/0019793916679601

[92] Figinski, T. F. (2017). Do employers favor those with military experience in the U.S. reserve forces? Evidence from a field experiment. *Defence and Peace Economics, 30*(2), 213–226. https://doi.org/10.1080/10242694.2017.1357521

[93] Flake, D. F. (2018). Do ban-the-box laws really work. *Iowa L. Rev., 104*(3), 1079. https://ilr.law.uiowa.edu/print/volume-104-issue-3/do-ban-the-box-laws-really-work

[94] Gaddis, S. M. (2014). Discrimination in the credential society: An audit study of race and college selectivity in the labor market. *Social Forces, 93*(4), 1451–1479. https://doi.org/10.1093/sf/sou111

[95] Galarza, F. B., & Yamada, G. (2017). Triple penalty in employment access: The role of beauty, race, and sex. *Journal of Applied Economics, 20*(1), 29–47. https://doi.org/10.1016/s1514-0326(17)30002-8

[96] Gerhards, J., Sawert, T., & Tuppat, J. (2020). Reversing the symbolic order of discrimination: Results from a field experiment on the discrimination of migrants and transgender people in theatre. *Journal of Ethnic and Migration Studies, 47*(15), 3622–3640. https://doi.org/10.1080/1369183x.2020.1754771

[97] 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. https://doi.org/10.1093/esr/jcy055

[98] Gorsuch, M., & Rho, D. (2018). *Customer prejudice & salience: The effect of the 2016 election on employment discrimination* (SSRN Electronic Journal Working Paper No. 3184984). University of Chicago, Becker Friedman Institute for Economics. https://doi.org/10.2139/ssrn.3184984

[99] Granberg, M., Andersson, P. A., & Ahmed, A. (2020). Hiring discrimination against transgender people: Evidence from a field experiment. *Labour Economics, 65*, 101860. https://doi.org/10.1016/j.labeco.2020.101860

[100] Guul, T. S., Villadsen, A. R., & Wulff, J. N. (2019). Does good performance reduce bad behavior? Antecedents of ethnic employment discrimination in public organizations. *Public Administration Review, 79*(5), 666–674. https://doi.org/10.1111/puar.13094

[101] Hipes, C., Lucas, J., Phelan, J. C., & White, R. C. (2016). The stigma of mental illness in the labor market. *Social Science Research, 56*, 16–25. https://doi.org/10.1016/j.ssresearch.2015.12.001

[102] Hipp, L. (2019). Do hiring practices penalize women and benefit men for having children? Experimental evidence from Germany. *European Sociological Review, 36*(2), 250–264. https://doi.org/10.1093/esr/jcz056

[103] Horváth, G. (2020). The impact of marital status on job finding: A field experiment in the Chinese labor market. *The B.E. Journal of Economic Analysis & Policy, 20*(4). https://doi.org/10.1515/bejeap-2019-0364

[104] Hou, Y., Liu, C., & Crabtree, C. (2020). Anti-Muslim bias in the Chinese labor market. *Journal of Comparative Economics, 48*(2), 235–250. https://doi.org/10.1016/j.jce.2019.12.001

[105] Jackson, M. (2009). Disadvantaged through discrimination? The role of employers in social stratification1. *The British Journal of Sociology, 60*(4), 669–692. https://doi.org/10.1111/j.1468-4446.2009.01270.x

[106] Jacquemet, N., & Yannelis, C. (2012). Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market. *Labour Economics, 19*(6), 824–832. https://doi.org/10.1016/j.labeco.2012.08.004

[107] Jaeger, D. A., Nunley, J. M., Seals, A., & Wilbrandt, E. J. (2020). *The Demand for Interns.* (National Bureau of Economic Research [NBER] Working Paper No. 26729). National Bureau of Economic Research. https://doi.org/10.3386/w26729

[108] Jibuti, D. (2018). *Discrimination against workers with visible tattoos: Experimental evidence from Germany* (The Center for Economic Research and Graduate Education – Economics Institute, Prague [CERGE-EI] Working Paper Series No. 628). CERGE-EI. https://ideas.repec.org/p/cer/papers/wp628.html

[109] Kaas, L., & Manger, C. (2012). Ethnic discrimination in Germany's labour market: A field

experiment. *German Economic Review, 13*(1), 1–20. https://doi.org/10.1111/j.1468-0475.2011.00538.x

[110] Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened résumés. *Administrative Science Quarterly, 61*(3), 469–502. https://doi.org/10.1177/0001839216639577

[111] Kleykamp, M. (2009). A great place to start? *Armed Forces & Society, 35*(2), 266–285. https://doi.org/10.1177/0095327x07308631

[112] 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. https://doi.org/10.1080/01419870.2019.1654114

[113] Lahey, J. N. (2008). Age, women, and hiring. *Journal of Human Resources, 43*(1), 30–56. https://doi.org/10.3368/jhr.43.1.30

[114] Lambert, J. R., & Akinlade, E. Y. (2019). Immigrant stereotypes and differential screening. *Personnel Review, 49*(4), 921–938. https://doi.org/10.1108/pr-06-2018-0229

[115] Larsen, E. N., & Di Stasio, V. (2019). Pakistani in the UK and Norway: Different contexts, similar disadvantage. Results from a comparative field experiment on hiring discrimination. *Journal of Ethnic and Migration Studies, 47*(6), 1201–1221. https://doi.org/10.1080/1369183x.2019.1622777

[116] Leasure, P., & Andersen, T. S. (2019). Race, criminal records, and certificates of relief: An experimental pilot study. *Deviant Behavior, 41*(9), 1065–1083. https://doi.org/10.1080/01639625.2019.1596539

[117] Lee, H.-A., & Khalid, M. A. (2015). Discrimination of high degrees: Race and graduate hiring in Malaysia. *Journal of the Asia Pacific Economy, 21*(1), 53–76. https://doi.org/10.1080/13547860.2015.1055948

[118] López Bóo, F., Rossi, M. A., & Urzúa, S. S. (2013). The labor market return to an attractive face: Evidence from a field experiment. *Economics Letters, 118*(1), 170–172. https://doi.org/10.1016/j.econlet.2012.10.016

[119] Manant, M., Pajak, S., & Soulié, N. (2018). Can social media lead to labor market discrimination? Evidence from a field experiment. *Journal of Economics & Management Strategy, 28*(2), 225–246. https://doi.org/10.1111/jems.12291

[120] Maurer-Fazio, M. (2012). Ethnic discrimination in China's internet job board labor market. *IZA Journal of Migration, 1*(1). https://doi.org/10.1186/2193-9039-1-12

[121] Maurer-Fazio, M., & Lei, L. (2015). As rare as a panda. *International Journal of Manpower, 36*(1), 68–85. https://doi.org/10.1108/ijm-12-2014-0258

[122] Mavlikeeva, M., & Asanov, I. (2020, September 27–30). *Can group identity explain gender gap in recruitment process?* [Conference presentation]. VfS Annual Conference 2020 (Virtual Conference): Gender Economics, Berlin, Germany. https://ideas.repec.org/p/zbw/vfsc20/224647.html

[123] McGinnity, F., & Lunn, P. D. (2011). Measuring discrimination facing ethnic minority job applicants: An Irish experiment. *Work, Employment and Society, 25*(4), 693–708. https://doi.org/10.1177/0950017011419722

[124] Midtbøen, A. H. (2014). Discrimination of the second generation: evidence from a field experiment in Norway. *Journal of International Migration and Integration, 17*(1), 253–272. https://doi.org/10.1007/s12134-014-0406-9

[125] Mishel, E. (2016). Discrimination against queer women in the U.S. workforce. *Socius: Sociological Research for a Dynamic World*, *2*, 237802311562131. https://doi.org/10.1177/2378023115621316

[126] Mobasseri, S. (2019). Race, place, and crime: How violent crime events affect employment discrimination. *American Journal of Sociology, 125*(1), 63–104. https://doi.org/10.1086/703883

[127] Neumark, D., Burn, I., & Button, P. (2016). Experimental age discrimination evidence and the Heckman critique. *American Economic Review, 106*(5), 303–308. https://doi.org/10.1257/aer.p20161008

[128] Neumark, D., Burn, I., Button, P., & Chehras, N. (2019). Do state laws protecting older workers from discrimination reduce age discrimination in hiring? Evidence from a field experiment. *The Journal of Law and Economics, 62*(2), 373–402. https://doi.org/10.1086/704008

[129] Nunley, J. M., Pugh, A., Romero, N., & Seals, R. A. (2015). Racial discrimination in the labor market for recent college graduates: Evidence from a field experiment. *The B.E. Journal of Economic Analysis & Policy, 15*(3), 1093–1125. https://doi.org/10.1515/bejeap-2014-0082

[130] Oreopoulos, P. (2011). Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. *American Economic Journal: Economic Policy, 3*(4), 148–171. https://doi.org/10.1257/pol.3.4.148

[131] Oreopoulos, P., & Dechief, D. (2012). *Why do some employers prefer to interview Matthew, but not Samir? New evidence from Toronto, Montreal, and Vancouver* (SSRN Electronic Journal Working Paper No. 2018047). University of Chicago, Becker Friedman Institute for Economics. https://doi.org/10.2139/ssrn.2018047

[132] Patacchini, E., Ragusa, G., & Zenou, Y. (2014). Unexplored dimensions of discrimination in Europe: Homosexuality and physical appearance. *Journal of Population Economics, 28*(4), 1045–1073. https://doi.org/10.1007/s00148-014-0533-9

[133] 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. https://doi.org/10.1016/j.labeco.2006.01.006

[134] Pierné, G. (2013). Hiring discrimination based on national origin and religious closeness: Results from a field experiment in the Paris area. *IZA Journal of Labor Economics, 2*(1). https://doi.org/10.1186/2193-8997-2-4

[135] Ramos, M., Thijssen, L., & Coenders, M. (2019). Labour market discrimination against Moroccan minorities in the Netherlands and Spain: A cross-national and cross-regional comparison. *Journal of Ethnic and Migration Studies, 47*(6), 1261–1284. https://doi.org/10.1080/1369183x.2019.1622824

[136] Riach, P. A. (2015). A field experiment investigating age discrimination in four European labour markets. *International Review of Applied Economics, 29*(5), 608–619. https://doi.org/10.1080/02692171.2015.1021667

[137] Riach, P. A., & Rich, J. (2006). An experimental investigation of sexual discrimination in hiring in the English labor market. *The B.E. Journal of Economic Analysis & Policy, 6*(2). https://doi.org/10.2202/1538-0637.1416

[138] Rivera, L. A., & Tilcsik, A. (2016). Class advantage, commitment penalty. *American Sociological Review, 81*(6), 1097–1131. https://doi.org/10.1177/0003122416668154

[139] Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring. *Journal of Human Resources, 44*(3), 710–735. https://doi.org/10.3368/jhr.44.3.710

[140] Ruffle, B. J., & Shtudiner, Z. (2015). Are good-looking people more employable? *Management Science, 61*(8), 1760–1776. https://doi.org/10.1287/mnsc.2014.1927

[141] Saeed, A., Maqsood, S., & Rafique, A. (2019). Color matters: Field experiment to explore the impact of facial complexion in Pakistani labor market. *Journal of the Asia Pacific*

Economy, 24(3), 347–363. https://doi.org/10.1080/13547860.2019.1612974

[142] Siddique, Z. (2011). Evidence on caste based discrimination. *Labour Economics, 18*, S146–S159. https://doi.org/10.1016/j.labeco.2011.07.002

[143] Spencer, N., Urquhart, M.-A., & Whitely, P. (2019). Class discrimination? Evidence from Jamaica: A racially homogeneous labor market. *Review of Radical Political Economics, 52*(1), 77–95. https://doi.org/10.1177/0486613419832674

[144] Stone, A., & Wright, T. (2013). When your face doesn't fit: Employment discrimination against people with facial disfigurements. *Journal of Applied Social Psychology, 43*(3), 515–526. https://doi.org/10.1111/j.1559-1816.2013.01032.x

[145] Thijssen, L., Coenders, M., & Lancee, B. (2021a). Ethnic discrimination in the Dutch labor market: Differences between ethnic minority groups and the role of personal information about job applicants—evidence from a field experiment. *Journal of International Migration and Integration, 22*(3), 1125–1150. https://doi.org/10.1007/s12134-020-00795-w

[146] Thijssen, L., Lancee, B., Veit, S., & Yemane, R. (2021b). Discrimination against Turkish minorities in Germany and the Netherlands: Field experimental evidence on the effect of diagnostic information on labour market outcomes. *Journal of Ethnic and Migration Studies, 47*(6), 1222–1239. https://doi.org/10.1080/1369183x.2019.1622793

[147] Thomas, K. (2018). The labor market value of taste: An experimental study of class bias in U.S. employment. *Sociological Science, 5*, 562–595. https://doi.org/10.15195/v5.a24

[148] Tilcsik, A. (2011). Pride and prejudice: Employment discrimination against openly gay men in the United States. *American Journal of Sociology, 117*(2), 586–626. https://doi.org/10.1086/661653

[149] Tinsley, M. (2012). *Too much to lose: Understanding and supporting Britain's older workers.* Policy Exchange. https://policyexchange.org.uk/publication/too-much-to-lose-understanding-and-supporting-britains-older-workers

[150] Tunstall, R., Green, A., Lupton, R., Watmough, S., & Bates, K. (2013). Does poor neighbourhood reputation create a neighbourhood effect on employment? The results of a field experiment in the UK. *Urban Studies, 51*(4), 763–780. https://doi.org/10.1177/0042098013492230

[151] Uggen, C., Vuolo, M., Lageson, S., Ruhland, E., & K. Whitham, H. (2014). The edge of stigma: An experimental audit of the effects of low-level criminal records on employment.

Criminology, 52(4), 627–654. https://doi.org/10.1111/1745-9125.12051

[152] Valfort, M.-A. (2020). Anti-Muslim discrimination in France: Evidence from a field experiment. *World Development, 135*, 105022. https://doi.org/10.1016/j.worlddev.2020.105022

[153] Van den Berg, C. J. W., Blommaert, L., Bijleveld, C. C. J. H., & Ruiter, S. (2020). Employment opportunities for ex-offenders. *Research in Social Stratification and Mobility, 65*, 100476. https://doi.org/10.1016/j.rssm.2020.100476

[154] Veit, S., & Thijsen, L. (2019). 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. https://doi.org/10.1080/1369183x.2019.1622825

[155] Vernby, K., & Dancygier, R. (2019). Can immigrants counteract employer discrimination? A factorial field experiment reveals the immutability of ethnic hierarchies. *PLOS ONE, 14*(7), e0218044. https://doi.org/10.1371/journal.pone.0218044

[156] Villadsen, A. R., & Wulff, J. N. (2018). Is the public sector a fairer employer? Ethnic employment discrimination in the public and private sectors. *Academy of Management Discoveries, 4*(4), 429–448. https://doi.org/10.5465/amd.2016.0029

[157] Vuolo, M., Uggen, C., & Lageson, S. (2017). Race, recession, and social closure in the low-wage labor market: Experimental and observational evidence. In S. Vallas (Ed.), *Emerging Conceptions of Work, Management and the Labor Market: Research in the Sociology of Work* (Vol. 30, pp. 141–183). Emerald Publishing Limited. https://doi.org/10.1108/s0277-283320170000030007

[158] Wallace, M., Wright, B. R. E., & Hyde, A. (2014). Religious affiliation and hiring discrimination in the American South. *Social Currents, 1*(2), 189–207. https://doi.org/10.1177/2329496514524541

[159] Weichselbaumer, D. (2014). Testing for discrimination against lesbians of different marital status: A field experiment. *Industrial Relations: A Journal of Economy and Society, 54*(1), 131–161. https://doi.org/10.1111/irel.12079

[160] Weichselbaumer, D. (2017). Discrimination against migrant job applicants in Austria: An experimental study. *German Economic Review, 18*(2), 237–265. https://doi.org/10.1111/geer.12104

[161] Weichselbaumer, D. (2019). Multiple discrimination against female immigrants wearing

headscarves. *ILR Review, 73*(3), 600–627. https://doi.org/10.1177/0019793919875707

[162] Widner, D., & Chicoine, S. (2011). It's all in the name: Employment discrimination against Arab Americans. *Sociological Forum, 26*(4), 806–823. https://doi.org/10.1111/j.1573-7861.2011.01285.x

[163] Wood, M., Hales, J., Purdon, S., Sejersen, T., & Hayllar, O. (2009*). A test for racial discrimination in recruitment practice in British cities* (Department for Work and Pensions Research Report No. 607). Department for Work and Pensions, United Kingdom. http://www.bollettinoadapt.it/old/files/document/3626ATESTFORRACIALDI.pdf

[164] Wu, K. (2017). *Beauty and Employment: A field experiment on appearance discrimination in China's labor market* (SSRN Electronic Journal Working Paper No. 3044200). University of Chicago, Becker Friedman Institute for Economics. https://doi.org/10.2139/ssrn.3044200

[165] Yemane, R. (2020). Cumulative disadvantage? The role of race compared to ethnicity, religion, and non-white phenotype in explaining hiring discrimination in the U.S. labour market. *Research in Social Stratification and Mobility, 69*, 100552. https://doi.org/10.1016/j.rssm.2020.100552

[166] Yemane, R., & Fernández-Reino, M. (2019). Latinos in the United States and in Spain: The impact of ethnic group stereotypes on labour market outcomes. *Journal of Ethnic and Migration Studies, 47*(6), 1240–1260. https://doi.org/10.1080/1369183x.2019.1622806

[167] Zhou, X., Zhang, J., & Song, X. (2013). *Gender discrimination in hiring: Evidence from 19,130 resumes in China* (SSRN Electronic Journal Working Paper No. 2195840). University of Chicago, Becker Friedman Institute for Economics. https://doi.org/10.2139/ssrn.2195840

[168] Zschirnt, E. (2020). Evidence of Hiring Discrimination Against the Second Generation: Results from a Correspondence Test in the Swiss Labour Market. *Journal of International Migration and Integration, 21*(2), 563–585. https://doi.org/10.1007/s12134-019-00664-1

[169] Zschirnt, E., & Fibbi, R. (2019). *Do Swiss citizens of immigrant origin face hiring discrimination in the labour market?* (National Center of Competence in Research [NCCR] Working Paper No. 19). National Center of Competence in Research – The Migration-Mobility Nexus. https://nccr-onthemove.ch/wp_live14/wp-content/uploads/2019/02/nccrotm-WP20-Zschirnt-Fibbi_Feb19.pdf