

Do Some Countries Discriminate More than Others? Evidence from 97 Field Experiments of Racial Discrimination in Hiring

Lincoln Quillian,^a Anthony Heath,^b Devah Pager,^c Arnfinn H. Midtbøen,^d Fenella Fleischmann,^e Ole Hexel^{a,f}

a) Northwestern University; b) Nuffield College; c) Harvard University; d) Institute for Social Research, Oslo, Norway;

e) Utrecht University; f) Sciences Po, Paris, France

Abstract: Comparing levels of discrimination across countries can provide a window into large-scale social and political factors often described as the root of discrimination. Because of difficulties in measurement, however, little is established about variation in hiring discrimination across countries. We address this gap through a formal meta-analysis of 97 field experiments of discrimination incorporating more than 200,000 job applications in nine countries in Europe and North America. We find significant discrimination against nonwhite natives in all countries in our analysis; discrimination against white immigrants is present but low. However, discrimination rates vary strongly by country: In high-discrimination countries, white natives receive nearly twice the callbacks of nonwhites; in low-discrimination countries, white natives receive about 25 percent more. France has the highest discrimination rates, followed by Sweden. We find smaller differences among Great Britain, Canada, Belgium, the Netherlands, Norway, the United States, and Germany. These findings challenge several conventional macro-level theories of discrimination.

Keywords: discrimination; race; ethnicity; hiring; field experiments

Citation: Quillian, Lincoln, Anthony Heath, Devah Pager, Arnfinn H. Midtbøen, Fenella Fleischmann, and Ole Hexel. 2019. "Do Some Countries Discriminate More than Others? Evidence from 97 Field Experiments of Racial Discrimination in Hiring." *Sociological Science* 6: 467-496.

Received: March 7, 2019

Accepted: April 23, 2019

Published: June 17, 2019

Editor(s): Jesper Sørensen, Olav Sørensen

DOI: 10.15195/v6.a18

Copyright: © 2019 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

RACIAL and ethnic inequality represents a pervasive feature of modern societies. Particularly in the labor market, gaps between minority members and native whites appear large and persistent. The Organisation for Economic Co-operation and Development (OECD) (2015), for example, notes that unemployment rates among native-born children of immigrants (aged 15 to 34) were around twice as high as those of their peers from the white majority group in Belgium, France, Germany, the Netherlands, Norway, Sweden, and the United Kingdom; these patterns are strikingly similar to the disparities in unemployment rates between African Americans and whites in the United States (Austin 2013).

To some, these gaps are simply transitory frictions on the pathway toward integration and assimilation. Particularly in European countries that have experienced high rates of immigration, many expect that first-generation disadvantages will give way to subsequent generations that enjoy the full scope of citizenship (Jonsson, Kalter, and Tubergen 2018). In the United States, discrimination is similarly minimized by explanations for contemporary racial inequality that emphasize vestiges of historical experiences rather than contemporary barriers (Heckman 1998; Wilson 2012). By contrast, others point to persistent discrimination as a fundamental cause of contemporary racial-ethnic inequality (Feagin and Sikes 1994; Sidanius and

Pratto 2001). The extent to which active discrimination shapes the opportunities available to racial and ethnic minorities thus remains highly contested.

Despite similarities in these debates across countries, there are reasons to believe that the extent of discrimination may vary considerably by national context. Countries differ in their racial and immigration histories, current economic and social contexts, and public policies (Alba and Foner 2015). Although these conditions vary somewhat within national boundaries, many aspects of history, culture, and policy are structured primarily at the country level. However, little is established about how levels of labor market discrimination vary across countries and which minority groups are affected. Establishing national differences in discrimination is a prerequisite to better understanding the large-scale social, cultural, and policy factors that influence discrimination.

In this article, we aim to establish new facts about discrimination by providing estimates of the magnitude of hiring discrimination across major racial and ethnic groups in several countries in Europe and North America. To do this, we combine evidence from 97 field experiments of hiring, grounded in more than 200,000 job applications, to draw conclusions about discrimination across nine countries (for which we have at least three field experiments of hiring discrimination each): Belgium, Canada, France, Germany, Great Britain, the Netherlands, Norway, Sweden, and the United States. Field experiments have the advantage of strong causal validity in establishing discrimination, offering less biased estimators of discrimination relative to the predominate approach in prior studies based on residual gaps from statistical models (National Research Council 2004). We combine data across studies using techniques of meta-analysis, the branch of statistics concerned with combining results from multiple studies.

Background

As a framework for this cross-national research, we begin by considering why theories can be taken to support either relative similarity in levels of discrimination among countries in Western Europe and North America or, alternatively, support significant national differences in discrimination patterns. Developed Western countries share enough similarities in their racial histories, current economic and social contexts, and policies that we might expect a common matrix of discrimination against minority groups. But important differences along these dimensions exist as well, and in the absence of strong theory about the causes of discrimination, it is plausible that discrimination could be either fairly uniform or quite different across countries.

Modern racial divisions and prejudice have their historical basis in the ideologies developed as part of early group contact, especially justifications of the international slave trade and colonialism (Fredrickson 2002). Racist ideologies were elaborated as an intellectual scheme based on ideas of heredity by European biologists and later mixed with ideas of Darwinian evolution (Gould 1996). The result was a set of beliefs, ideas, and prejudices about the inferiority of nonwhite racial groups that were often quite similar across Western countries (Winant 2001).

In more recent times, the position of Western nations in the world system of migration and strong cultural links among Western countries suggest fairly similar responses to migration. Indeed, migration from the global South has sparked backlash in many Western countries (Semyonov, Raijman, and Gorodzeisky 2006; Golder 2016), as is evident in the rise of populist anti-immigrant parties in Europe and the recent election of Donald Trump in the United States. Views toward immigrants have also been similarly influenced by terror attacks tied to Islamic extremism, notably the attacks of September 11, 2001, in the United States and subsequent attacks in several European countries (Branton et al. 2011; Legewie 2013). Finally, ethnic minorities in Europe and North America face many similar problems, in particular unemployment (Heath and Cheung 2007). As noted previously, nonwhites in Europe and blacks in North America tend to have unemployment rates about twice the white native rate (OECD 2015).

Also similar are many elements in North American and European countries' legislation and practices regarding race and ethnicity. Countries have tended to copy legislation and practices from other countries, reflecting strong organizational isomorphism across countries (Meyer et al. 1997). A fairly similar set of antidiscrimination laws were adopted in North America and many Western European countries from the 1960s to the 1990s. In 2000, the European Union passed a series of race directives that mandated a range of antidiscrimination measures to be adopted by all member states, putting their legislative frameworks on racial discrimination on highly similar footing (European Union Agency for Fundamental Rights 2008).

On the other hand, these points of commonality are also accompanied by notable national differences in race and hiring practices. On the historical dimension, although European and North American countries were all influenced by European colonialism and the slave trade, they did not participate equally. The United States is unique among countries we examine in having a large resident population descended from enslaved persons. Some studies find evidence of connections between involvement with slavery and modern racial inequality (O'Connell 2012). And although some European countries have extensive colonial histories, such as England, France, and Belgium, countries such as Sweden and Norway had less involvement in colonialism.¹ Finally, although countries around the world were influenced by the civil rights movement, it was centered and most influential in the United States.

National differences in situational factors that influence discrimination also suggest that discrimination may differ significantly across countries. Group threat theory, for instance, suggests prejudice and discrimination as a reaction to threats triggered by minority group size and recent increases in minority group size, factors that differ across countries (Blalock 1967; Taylor 1998; Schlueter and Scheepers 2010). Poor economic conditions may also heighten feelings of threat and increase discrimination by majority against minority groups in ways that differ with the position of national economies (Quillian 1995).

Finally, countries differ in institutions relating to race and hiring. On race, some countries extensively measure and monitor racial and ethnic gaps, the most extensive being the United States. Others rarely monitor them or even bar the measurement or use of race as a category for many official purposes; France is the

most prominent example of a country that bars measurement of race and ethnicity (Beaman 2017; Simon 2008). The effects of these policies are debated, with some suggesting that recognition reifies race or ethnicity and increases discrimination (e.g., American Anthropological Association 1997), whereas others argue that recognition is critical for antidiscrimination enforcement (e.g., American Sociological Association 2002; Simon 2008). Hiring, too, varies in potentially important ways across countries. In European countries, it is more common to include a photograph at first application than in the United States, and in German-speaking countries, it is common for a detailed range of information, such as high school grades and reports of apprenticeships, to be submitted at first application (Weichselbaumer 2016). The United States is distinctive in having institutionalized practices encouraging diversity in hiring at many large corporations, including mandatory reporting of gender and racial-ethnic composition of employees to the government and forms of affirmative action (Dobbin 2011). Evidence indicates the effects of these policies on discrimination are inconsistent, but parts of antidiscrimination law and practice do reduce discrimination (Kalev, Dobbin, and Kelly 2006).

Measuring National Discrimination Levels

Little is established about how discrimination varies across national contexts, largely because of difficulties in measurement. Past studies aiming to assess discrimination levels across countries have generally relied on either indirect methods based on racial gaps or proxy reports. Both methods have serious shortcomings.

The most common method for assessing discrimination is often referred to as the residual method, named for its reliance on the residual from a statistical model that aims to control for all observable factors that may differentiate majority and minority group members, such as age, education, and work experience. The unexplained gap, or residual, is often interpreted as the effect of discrimination. Of course, many other unobserved factors may also contribute to the size of the residual in such equations, leading researchers to overestimate (and sometimes underestimate) the true effects of discrimination (National Research Council 2004; Quillian 2006).

A second method relies on self-reports from potential targets of discrimination, often gathered through surveys. Although self-reports of discrimination are quite common, it is difficult to adequately map perceptions of discrimination with actual discriminatory behavior. Given the subtle and often covert nature of contemporary discrimination (Bonilla-Silva 2006), targets of discrimination are often unaware that a discriminatory incident has taken place. On the other hand, some individuals may misperceive generic hostility or poor service for an act with discriminatory intent. Reconciling the disconnect between perceptions and behavior is no easy task and one that likely varies across contexts (National Research Council 2004).

A third method uses the frequency of formal complaints of discrimination or lawsuits alleging discrimination. This method, too, only captures discrimination that victims are aware of, and formal complaints or lawsuits are strongly influenced by institutional factors that discourage or encourage official grievance procedures (Pager 2007). Finally, reports from perpetrators of discrimination can also be used.

Such reports face the obvious problem that perpetrators are likely to underreport their discrimination, making this an unreliable method as well (Pager and Quillian 2005; Pager 2007). All of these methods have serious shortcomings that undercut the ability to clearly understand their results as measuring discrimination unambiguously and limit our ability to reliably compare patterns of discrimination across national contexts.

A method with better causal (internal) validity that has seen increasing use over the last 15 years is the field experimental method. Field experiments of hiring discrimination are experimental or quasi-experimental studies in which fictionalized candidates from different racial or ethnic groups apply for jobs. These include both resume audit studies, in which fictionalized resumes are submitted by mail, by e-mail, or through a website (e.g., Bertrand and Mullainathan 2004), and in-person audit studies, in which ethnically dissimilar but otherwise matched pairs of trained testers apply for jobs (e.g., Pager, Bonikowski, and Western 2009).

As the National Research Council (2004) points out, the problem of measuring discrimination is fundamentally a problem of causal inference: Discrimination measurement involves determining if and how much behavior is directly shaped by racial cues. It is for this reason that experimental methods, the gold standard for causal inference, provide robust methods to assess discrimination. For causal identification, field experiments of hiring discrimination rely on either matched pairs or randomization. In matched pairs, pairs of trained testers who are racially different apply for jobs. The applicants are matched to make them as similar as possible in job-relevant characteristics so that race and/or ethnicity is the only distinguishing characteristic. This includes giving applicants resumes with similar levels of qualifications, matching appearance, and training applicants to behave similarly. In studies using randomization, clues indicating race or ethnicity (such as racially or ethnically typed names) are randomly assigned to resumes, allowing randomization to equate majority and minority groups on these attributes.²

Field experiments have strong causal (internal) validity, but their reliability can be more of a problem. The number of applications in a single study is often fairly small, especially for in-person audits, and the magnitude of racial differences in outcomes from single studies is often not very precisely estimated. All but a tiny handful of field experimental studies are only conducted in a single country, making their results unsuitable for cross-national comparisons.³

To address the challenge of making national comparisons, we combine the results of many past field experimental studies through a systematic meta-analysis. Meta-analysis offers the opportunity to capitalize on the unique strength of the field experimental method—its high internal validity—while overcoming the limitations of single studies by providing a pooled estimate of hiring discrimination across multiple studies (Borenstein et al. 2009).

A number of other recent articles provide narrative reviews of field experiments of discrimination (e.g., Riach and Rich 2002; Pager and Shepherd 2008; Heath, Liebig, and Simon 2013; Bertrand and Duflo 2016; Baert 2018; Neumark 2018), but most focus on one country or a few countries in which similar field experiments have been performed, and none of these perform a formal meta-analysis to systematically combine evidence from different studies. We know of only one other

cross-national meta-analysis of field experimental studies: a notable recent article by Zschirnt and Ruedin (2016). For purposes of our analysis here, the most important difference from our study is that they do not attempt to estimate country-level differences. They include an aggregated comparison of Europe to North America and of German-speaking countries to non-German-speaking countries. This treats Europe and German-speaking countries as single units. As we discuss below, we find large differences in discrimination levels among European and North American countries.⁴

Procedure

Field experiments have been a method of growing popularity, with the result being that a large body of such studies now provide estimates of the levels of discrimination against racial and ethnic groups within many countries. For the large majority of field experimental studies of hiring, the primary outcome is a callback (a request to a candidate to return for an interview or a request for more information), indicating interest by the employer. Only a handful of audits follow the application process all the way through to the final hiring decision, too few to support comparison across countries.

Meta-analysis is a method widely used in medicine and psychology to aggregate the results of experimental studies through a secondary statistical analysis of their results (see Borenstein et al. 2009; Cooper, Hedges, and Valentine 2009). Using the method of meta-regression, with a database including all available field experimental studies of racial and ethnic discrimination in hiring, we model the relative rate of discrimination against minority groups as a function of the country, the minority group, and other study characteristics.

Our procedure follows three basic stages: first, to identify all existing field experiments of hiring discrimination; second, to develop a coding rubric and to code studies to produce a database of their results; and third, performing the statistical meta-analysis to draw conclusions from the combined results. We discuss each of these steps in turn (some of our discussion in this section reprises parts of Quillian et al. 2017).

Identifying Relevant Studies

We aimed to include in our meta-analysis all existing studies, published or unpublished, that use a field experimental method and that provide contrasts in hiring-related outcomes between different racial and ethnic groups in North America and Europe (to the end of 2016, when we completed data collection). This includes both in-person audit studies and resume studies. We also required that contrasts between racial or ethnic groups in included studies were made between fictionalized applicants that were, on average, equivalent in their labor market-relevant characteristics—for instance, that the majority and minority applicants have similar levels of education and work experience—because otherwise, discrimination estimates are confounded with the difference in nonracial characteristics.

We used three methods to identify relevant field experiments: searches in bibliographic databases, citation searches, and an e-mail request to corresponding authors of field experiments of racial-ethnic discrimination in labor markets and other experts on field experiments and discrimination.

Our first search method was bibliographic search. Our search covered the following bibliographic databases and working paper repositories: Thomson's Web of Science (Social Science Citation Index), ProQuest Sociological Abstracts, ProQuest Dissertations and Theses, LexisNexis, Google Scholar, and National Bureau of Economic Research Working Papers. We searched for some combination of "field experiment," "audit study," or "correspondence study" and sometimes included the term "discrimination," with some variation depending on the search functions of the database. To improve our coverage of non-English publications, we also searched two French-language indexes, Cairn.info and Persée, and two international sources, IZA Discussion Papers (a German working paper archive) and International Labour Organization International Migration Papers. Finally, we conducted a search with Italian, Spanish, Portuguese, and Dutch translations of the search terms and other terms frequently used in these languages to describe field experiments in hiring discrimination in Google Scholar. The search was first performed in March 2014 and repeated in August and September 2014 and in November 2015. Searches in Italian, Spanish, Portuguese, and Dutch were conducted in November 2015 and February 2016.

Our second search method was citation search. Working from the initial set of studies located through bibliographic search, we examined the bibliographies of all review articles and eligible audit studies to find additional field experiments of hiring discrimination.

Our last search method was an e-mail request to authors of existing field experiments of discrimination. From our list of audit studies identified by bibliographic and citation search, we compiled a list of e-mail addresses of authors of existing field experiments of discrimination. To this, we added the addresses of several well-known experts on field experiments, notably authors of literature review articles about field experiments. Our e-mail request asked for citations or copies of field discrimination studies that were published, unpublished, or ongoing. We also asked that authors refer us to any other researchers who may have recent or ongoing field experiments.

The e-mail requests were conducted in two phases. In the initial wave, 131 apparently valid e-mail addresses were contacted. We received 56 responses. We also sent out a second wave of 68 e-mails, which consisted of additional authors identified from the initial wave of surveys and some corrected e-mail addresses. We received 19 responses to this second wave of e-mail surveys.

Overall, our search located more than 100 studies and included contrasts between white and minority groups who were, on average, equivalent in their labor market-relevant characteristics (e.g., education, experience level in the labor market, etc.) and who otherwise met our inclusion criteria.

Table 1: Number of effects and studies by country and minority groups.

Country	Effects	Studies	Applications	Minority Groups with Effect Sizes (Study Term)
Belgium	6	3	5,989	Congolese, Italian, Moroccan, Turkish
Canada	18	7	28,733	African, Arab, Black, Chinese, Greek, Indian, Indo-Pakistani, Latino, Middle Eastern, West Indian, White Immigrant
France	28	21	37,810	African, Antillean, Asian, Franco-North African, Moroccan, North African, Senegalese, Sub-Saharan African, Vietnamese
Germany	5	5	8,856	Turkish
Great Britain	32	10	7,887	African, Asian (South Asian), Australian, Black African, Black Caribbean, Chinese, Cypriot, French, Greek, Indian, Italian, Pakistani, Pakistani/Bangladeshi, West Indian
Netherlands	19	10	8,012	Antillean, Arab, Black Surinamer, Hindustani, Moroccan, Spanish, Surinamese, Turkish
Norway	4	4	3,582	Pakistani
Sweden	8	7	26,119	Arab, Middle Eastern
United States	39	30	73,024	African American, Arab American, Asian, Black, Hispanic, Latino, Somali, White Jewish
Total	159	97	200,012	
Minority Group (Short Name)				Notes
African/Black (Black)	58	52		Excludes North African
European/White Immigrant (White)	13	11		
Middle Eastern/North African (MENA)	47	42		Includes Turkish
Latin American/Hispanic (Hispanic)	10	10		
Asian (Asian)	31	22		Includes East Asian and South Asian
Total	159			

Note: For the minority group tabulation, a study can be present in more than one racial or ethnic category. Effects are estimates of discrimination against minority groups. Some studies have estimates for multiple minority groups.

Sample and Selection of Countries

Our selection of the nine countries we focus on was driven by data considerations. We found that standard errors were too large to be informative for country estimates based on only one or two studies. There are nine countries in our sample with three or more field experiments: Belgium, Canada, France, Germany, Great Britain, the Netherlands, Norway, Sweden, and the United States.

Our final sample in these nine countries includes a total of 97 distinct studies, including 159 estimates of discrimination against distinct minority groups. Many studies include more than one minority group (e.g., African American and Hispanic or Pakistani and Chinese people), which is why there are more effects than studies. Table 1 shows the breakdown of effects and studies across the nine countries. It also lists specific minority groups. For analysis, we recoded the specific minority groups into five broader racial and/or ethnic target groups (brief name is in parentheses): African/black (black), European/white (white), Middle Eastern/North African (MENA), Latin American/Hispanic (Hispanic), and Asian (Asian). African/black includes persons descended from populations in all parts of Africa except North Africa. Asian includes persons descended from peoples of East and South Asia but predominately South Asian in our data.⁵

Coding

We coded key characteristics of the studies into a database for our analysis. Coding was based on a coding rubric, which listed the characteristics and included coding instructions. To develop the rubric, we initially read several studies, and based on this, we developed an initial coding rubric of factors we thought might influence measured rates of discrimination. It was subsequently refined as coding progressed.

To ensure reliability, almost all studies in our analysis were coded independently by two raters. Studies were coded by fluent readers in English, French, German, and Dutch. One study report in Swedish and one in Norwegian were coded by a single rater.

We then reconciled the results of the two codings, performing further investigations to find the correct answer on coding decisions in cases of disagreement. The variables coded were factual in nature (e.g., year of publication, counts of positive and negative responses for the white and nonwhite group, etc.); the main sources of disagreement in coding were difficulty in understanding the text or procedures of a particular study or occasional judgment calls about what “fit” in a particular category (e.g., what jobs are “blue collar”?).

The coding involved two levels of information: study level and effect level. Study-level characteristics are constant for the entire study, such as year of publication and nature of the published outlet of the study (e.g., country and year of the study). Effect estimates refer to estimates of discrimination against a group, with the number of effect sizes for a study depending on the number of target groups the study includes. A study that contrasts whites with African Americans and Latinos, for instance, would produce two effect sizes.

We coded effects that measure discrimination based on counts of callbacks by racial or ethnic group. Most studies included counts of outcomes in their research report. When the study did not include counts of outcomes in the research report, we requested counts from the authors and excluded the study if we did not receive the counts. We used all native white and minority testers in computing effect sizes except for the few cases in which the groups were nonequivalent in their labor market characteristics, most often when minorities were given somewhat stronger background qualifications than native whites.

Outcome: The Discrimination Ratio

Our basic outcome, the “effect” measure in meta-analysis terminology, is the ratio of the percentage of callbacks to job applications by members of the majority group to the minority group (often called the “relative risk” or “risk ratio”). Formally, if c^w is the number of callbacks received by white natives, c^m is the number of callbacks received by a minority race or ethnic group, n^w is the number of applications submitted by white native applicants, and n^m is the number of applications submitted by minority applicants, then the discrimination ratio is $(c^w/n^w)/(c^m/n^m)$. We calculated this ratio based on counts of the number of positive responses to the majority group and minority group from each study. Ratios greater than 1 indicate the majority received more positive responses than the minority, with the amount greater than 1 multiplied by 100 indicating the percentage more callbacks

for the majority group relative to the minority group. Numbers greater than 1 indicate higher discrimination against the minority group. Because studies equate the groups on their nonracial characteristics either through matching and assignment of job-relevant characteristics (audits) or through random assignment (many correspondence studies), no further controls are required for internally valid estimates of discrimination.⁶

The discrimination ratio may be interpreted as the number of applications that must be submitted by a minority applicant to expect an equal chance of a callback as a white applicant. With a discrimination ratio of 2.25, a minority candidate has to send out 2.25 applications for every application submitted by the white tester in order to expect to receive the same number of callbacks.

Two other measures that could be used instead of the discrimination ratio are the difference in proportions of positive responses and the odds ratio. We prefer the discrimination ratio to the difference in proportions because it is less sensitive to the base rate of the outcome (see Borenstein et al. 2009). For the difference in proportions, high-base-rate studies dominate low-base-rate studies in terms of the measure: For instance, a study in which 45 percent of whites and 40 percent of blacks receive a positive response gives the same discrimination difference estimate as one in which 9 percent of whites and 4 percent of blacks receive positive responses, although our view is the latter shows much higher discrimination than the former. Another potential choice with good statistical properties is the odds ratio, but we prefer the ratio of positive responses because it is much more easily interpretable. Online supplement Table S3 shows results of our basic model (corresponding to Table 3, model 2) using the odds ratio outcome. The basic results are similar to the discrimination ratio.

Meta-analysis Model

We employ a random-effects specification for the meta-analysis model (Raudenbush 2009). The random-effects specification incorporates in the error structure a variance component capturing variation in outcomes across studies because of factors that vary at the study level but are not controlled; the model is a type of multilevel model with random components at the applicant and effect levels. A random-effects specification is recommended whenever there is reason to believe that the effect estimated by the studies in a meta-analysis is likely to vary because of design features of studies that are not directly controlled for in the analysis rather than representing a single underlying effect that is constant across the whole population. This is the case in our analysis because we expect discrimination against a target group may depend on the country, the situation the study considers (e.g., the occupational categories), the (falsified) credentials of the applicants, the types of jobs applied for, and so on.

In practice, the study-level variance component (random effect) has the effect of inflating standard errors to reflect unaccounted for differences in outcomes among studies. Between-study variability is considerable, so this adjustment increases standard errors substantially. Standard errors from single studies do not incorporate

this, and for this reason, our inference procedures are arguably more conservative than those used in individual field experiments to calculate standard errors.

Meta-regression allows the rate of discrimination to be a function of a vector of characteristics of the studies and effects plus (in the random-effects specification) residual study-level heterogeneity (between-study variance not explained by the covariates).⁷ The model assumes the study-level heterogeneity follows a normal distribution around the linear predictor. If y_{ij} is the discrimination ratio for the j th effect size in the i th study, then the meta-analysis model is

$$\ln(y_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta} + u_i + e_{ij}, \text{ where } u_i \sim N(0, \tau^2) \text{ and } e_{ij} \sim N(0, \sigma_{ij}^2),$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients (including a constant), and $\mathbf{x}_{ij}$ is a $1 \times k$ vector of covariate values in study i and effect size j (k is the number of covariates including a 1 for a constant). Following standard practice in the meta-analysis literature, we log the response ratio to reduce the asymmetry of the ratio. Residual between-study variance is τ^2 , estimated as part of the meta-analysis model. Models were estimated by restricted maximum likelihood with the metafor package in R (Viechtbauer 2010).

In this study, we include covariates representing different countries (as dummies) and target groups (as dummies) as our primary independent variables. After the most basic model, we include additional covariates that may influence discrimination. These covariates and descriptive statistics are shown in Table 2. Two of the variables we use as predictors correspond directly to situational explanations discussed previously: the unemployment rate and the percentage of immigrants. These are measured at the region or metropolitan level; for studies with multiple regions or metropolitan areas, this is an average over fieldsites weighted by the percentage of study applicants at each fieldsite.

Study Weighting

In estimation, each observation is weighted by inverse variance; the variance of each observation is $\tau^2 + \sigma_{ij}^2$, so each observation is weighted by $1/(\tau^2 + \sigma_{ij}^2)$. The parameter τ^2 is between-study variance, which is estimated as a parameter in each meta-regression. The parameter σ_{ij}^2 is the variance of the log discrimination ratio of the j th effect size in the i th study, σ_{ij}^2 . In large part, this reflects the sample size (number of applications submitted), with larger variability (giving less weight) to effect sizes being based on small samples of applications.

We calculate σ_{ij}^2 from counts of applications and callbacks. For studies that are unpaired or do not report paired outcomes, the variance of the log discrimination ratio for the j th minority group in the i th study for callbacks is estimated by

$$\sigma_{ij}^2 = \text{Var}(\ln(y_{ij})) = 1/c_{ij}^w - 1/n_{ij}^w + 1/c_{ij}^m - 1/n_{ij}^m.$$

This is Borenstein et al.'s (2009) formula 5.3. The c and n terms are counts of callbacks and applications for white native (w) and minority (m) groups as defined in the discussion of the discrimination ratio above.

Table 2: Descriptive statistics, predictor variables.

Study Methods	Effects	Studies	
Resume Audit	134	80	
In-Person Audit	25	17	
Tester Gender			
Testers Male Only	46	38	
Testers Female Only	9	8	
Testers Both Male and Female	104	53	
Applicant Education (Most Common Level)			
High School or Less	42	31	
Some College or Post-Vocational Degree	34	25	
College or More	38	24	
Education Information Missing	45	17	
Occupational Controls			
Includes Blue-Collar Jobs (1 = Yes)	79	47	
Includes Jobs with Customer Contact (1 = Yes)	118	73	
Includes Jobs with an Office Focus (1 = Yes)	121	73	
Immigrant Status			
Minority Applicants Foreign Born (1 = Yes)	42	17	
Minority Applicants Have Foreign Nationality (1 = Yes)	11	7	
Minority Applicants Final Credential Foreign (1 = Yes)	9	2	
	Mean (Effect Level)	Standard Deviation	N (Effects/ Studies)
Year of Fieldwork	1999.7	14.7	159/97
Unemployment Rate of Local City/Region	6.9%	2.4%	158/97
Percentage Immigrants in Local City/Region	13.6%	10.9%	155/94

Note: Effects are distinct estimates of discrimination against minority groups. Some studies include estimates of discrimination against multiple minority groups.

For studies that use a paired design—with one minority and one white native applicant applying for each job—and report paired outcomes, we use an alternative formula because the pairing affects the variability of the ratio (see Zhou 2007). If p^a is the number of pairs in which both majority and minority testers receive a callback, p^b is the number of pairs in which the majority tester received a callback but not the minority tester; and if p^c is the number of pairs in which the minority tester received a callback but not the majority tester, then the variance of the log discrimination ratio for the j th minority group in the i th study with paired data is

$$\sigma_{ij}^2 = \text{Var}(\ln(y_{ij})) = \frac{p_{ij}^b + p_{ij}^c}{(p_{ij}^a + p_{ij}^b)(p_{ij}^a + p_{ij}^c)}.$$

For studies that are paired between the majority and minority but in which paired outcomes are not reported, we use formulas for the standard error of unpaired groups. This formula will slightly overestimate the standard error of the effect, underweighting these studies a bit in computing the overall effect and slightly inflating the overall cross-study standard error.

Adjusting Standard Errors for Clustering of Effects within Studies

Most studies in our analysis contribute more than one effect size (estimation of discrimination against a minority group). For instance, a study may have discrimination estimates against persons of African/black descent and Middle Eastern/North African descent, giving two effect sizes when contrasted with native whites. These two effect sizes are not independent. This is because first, in many cases, the contrast group of whites used to calculate the African/black and Middle Eastern/North African effect sizes is the same pool of fictitious applicants, creating dependence through a common control group. Second, both are derived from common study procedures, such as using the same set of nonethnic resume characteristics.

To adjust standard errors for this dependence, we use robust standard errors (command “robust”) as implemented in the *metafor* package in the R statistical language (Viechtbauer 2010). We also performed checks of some models using the *robumeta* package for meta-analysis with robust variance estimators with small-sample adjustments, which produced generally similar results (Fisher and Tipton 2015).

Results

We begin by examining descriptively how levels of discrimination vary by country and minority group. We use a random-effects meta-regression with distinct effects for minority group in each country—including country, target group, and interactions of country and target group—with no other controls. Predicted levels of the discrimination ratio for each country are shown in Figure 1. The dot is the point estimate of the country and target group average discrimination ratio, and the line is a 95 percent confidence interval. The number below each confidence interval is the number of studies used to compute the effect. The discrimination ratio is the ratio of callbacks for white natives to the indicated minority group. We focus on overall patterns in the figure rather than results of significance (or insignificance) of an individual country-by-group cell.⁸

The figure shows nearly ubiquitous discrimination against racial and ethnic minority groups: For all 25 of 26 target groups in the figure, point estimates of the discrimination ratio are greater than 1, indicating discrimination against minority groups (the one exception is white immigrants in the Netherlands, which is 0.95). There is no evidence of “reverse” discrimination against white natives. For several of the 26 target groups, the effects are not statistically significantly different from 1, but a more careful consideration suggests this largely reflects low power in estimating effects for some groups. There are 15 group discrimination estimates based on four or more field experimental studies, for which we have higher power. In 13 of these 15 estimates, a discrimination ratio of 1 (no discrimination) is outside of the 95 percent confidence interval, indicating statistically significant discrimination at $p < 0.05$ (two tailed). The two group estimates (based on four or more studies) that are not significant are for European/white immigrants (in Canada and Great Britain), suggesting lower discrimination against European/white immigrants than nonwhite groups. These results support the conclusion of ubiquitous discrimination

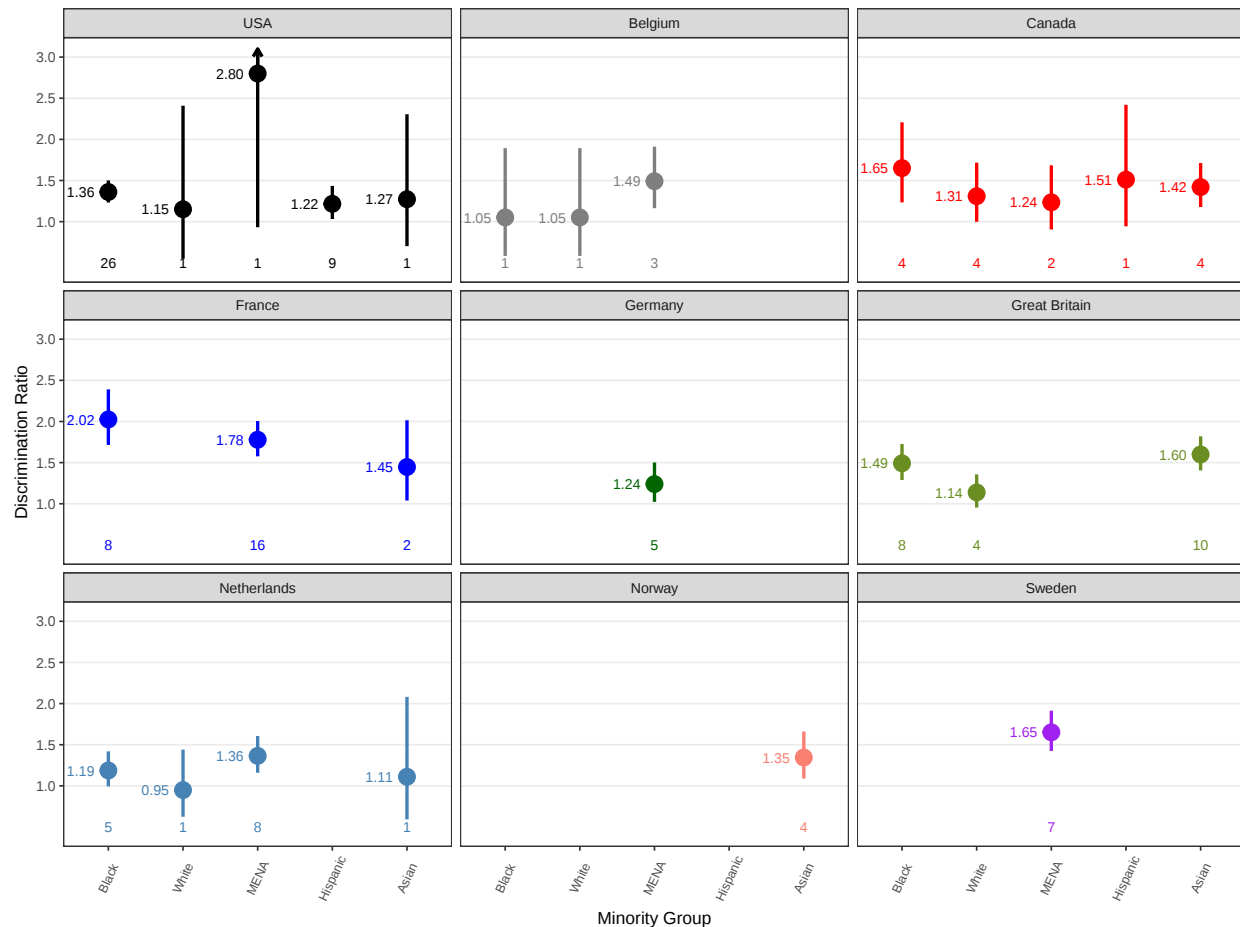


Figure 1: Average discrimination ratios by country and target group. Lines are 95 percent confidence intervals. The number left of the line is the point estimate. The number below the line is the number of studies used to compute effect.

against nonwhite groups. For white immigrants, by contrast, discrimination is lower and is often not statistically significant.

To further explore the sources of national and group differences and to account for other measured differences among field experimental studies, we build a meta-regression model of the discrimination ratio as a function of country, target group, and other factors. Standard errors are adjusted for correlated effects when there are multiple estimates of discrimination within the same study.

In our simplest model, we analyze the log discrimination ratio as an additive sum of a country and a target group effect. Table 3, model 1 shows these basic estimates. The United States is the reference group for the country dummy variables, and African/black is the reference for the group effects.

From model 1 of Table 3, two results stand out. First, France has the highest level of discrimination, with a discrimination ratio 33.6 percent ($\exp[0.29]-1$) higher than that in the United States.⁹ Second, immigrant groups from European-origin

Table 3: Meta-regression estimates of log discrimination ratio on country, minority, and controls.

	Country and Minority Group (1)	Base Controls (2)	Foreign Characteristics (3)	Contextual Variables (4)
Country (Reference = United States)				
Belgium	0.03 (0.15)	0.04 (0.12)	0.04 (0.13)	0.06 (0.14)
Canada	0.1 (0.80)	0.10 (0.11)	0.10 (0.12)	0.00 (0.16)
France	0.29 [†] (0.11)	0.36 [†] (0.1)	0.35 [†] (0.11)	0.39 [†] (0.14)
Germany	−0.09 (0.11)	−0.08 (0.13)	−0.11 (0.14)	−0.11 (0.13)
Great Britain	0.12 (0.08)	0.1 (0.14)	0.12 (0.13)	0.18 (0.15)
Netherlands	−0.06 (0.11)	0.03 (0.1)	0.02 (0.12)	0.08 (0.12)
Norway	−0.01 (0.1)	0.0 (0.12)	0.01 (0.13)	0.0 (0.14)
Sweden	0.19 (0.13)	0.27 [†] (0.15)	0.25 (0.16)	0.22 (0.15)
Target Group (Reference = African/Black)				
European Immigrant	−0.26 [†] (0.05)	−0.21 [†] (0.07)	−0.19 [†] (0.07)	−0.22 [†] (0.07)
Middle Eastern/North African	0.0 (0.09)	−0.01 (0.08)	0.0 (0.08)	0.03 (0.08)
Hispanic	−0.1 [†] (0.06)	−0.11 (0.08)	−0.11 (0.09)	−0.1 (0.09)
Asian	−0.01 (0.06)	0.04 (0.06)	0.04 (0.06)	0.02 (0.07)
Applicant Gender (Reference = Both)				
Testers Male Only (1 = Yes)		−0.03 (0.07)	−0.04 (0.08)	−0.02 (0.08)
Testers Female Only (1 = Yes)		−0.06 (0.08)	−0.07 (0.08)	−0.02 (0.08)
Most Common Level of Applicant Education (Reference = High School or More)				
Some College or Post-High School Vocational Degree		−0.08 (0.09)	−0.08 (0.1)	−0.04 (0.09)
College or More		−0.17 [*] (0.08)	−0.17 [†] (0.09)	−0.14 [†] (0.08)
Education Information Missing		−0.13 [†] (0.07)	−0.11 (0.09)	−0.11 (0.08)
Study Attributes				
In-Person Audit (1 = Yes)		0.11 [*] (0.05)	0.12 [†] (0.06)	0.11 [†] (0.06)
Year of Fieldwork (Four-Digit Year, Coefficient × 10)		0.01 (0.03)	0.0 (0.04)	0.01 (0.04)
Occupational Controls				
Includes Blue-Collar Jobs (1 = Yes)		−0.11 [†] (0.06)	−0.11 (0.07)	−0.11 (0.07)
Includes Jobs with Customer Contact (1 = Yes)		0.06 (0.07)	0.05 (0.08)	0.05 (0.07)
Includes Jobs with an Office Focus (1 = Yes)		0.09 (0.08)	0.1 (0.08)	0.07 (0.08)
Foreign Characteristics				
Minority Applicants Foreign Born (1 = Yes)			−0.06 (0.09)	
Minority Applicants Have Foreign Nationality (1 = Yes)			0.07 (0.01)	
Minority Applicants Have (Some) Foreign Credentials (1 = Yes)			−0.08 (0.13)	

Table 3 continued

	Country and Minority Group (1)	Base Controls (2)	Foreign Characteristics (3)	Contextual Variables (4)
Situational Variables				
Unemployment Rate in Local Area (Metropolitan/Regional)				0.01 (1.08)
Percentage Immigrant in Local Area				0.64 (0.56)
Intercept	0.31 [†] (0.05)	0.3* (0.13)	0.31* (0.14)	0.18 (0.16)
τ^2 (Between-Study Variable)	0.038	0.043	0.036	0.033
I^2 (% Between-Study Variable)	85	82.3	82.2	81
<i>N</i> Effects	159	159	159	155
<i>N</i> Studies	97	97	97	94

Standard errors are in parentheses. [†] $p < 0.01$; * $p < 0.05$; ‡ $p < 0.1$ (two-tailed tests).

countries experience significantly less discrimination than black or Sub-Saharan African persons. African/black, Middle Eastern/North African, and Asian minority groups all experience fairly similar levels of discrimination.

These models do not control for differences in characteristics of studies that may confound group and national differences. In model 2, we add controls for the gender of applicants, applicant education levels, whether the study is done in person or through the mail or Internet, year of fieldwork, and controls for occupational categories. Most of these controls are not significant, but we find stronger evidence of discrimination in face-to-face studies than resume studies. This may be because in-person applications provide stronger signals about ethnic identity than do names on resumes (see Gaddis 2017). Alternatively, the potential for unconscious bias on the part of actors who are aware of the study's purpose may be a factor (Heckman and Siegelman 1992). We also find less discrimination for jobs that require a college degree than jobs requiring only a high school degree (or national equivalent) or less. Our analysis does not permit us to determine why this is the case, but one possibility is that the material in the resumes of college-educated applicants tends to be more extensive and to contain more detail and thus reduces employers' uncertainty about applicants' characteristics.

The country differences are large compared with most other covariates in the model: They tend to be larger than minority group effects or most of the controls. The country effects are graphed in Figure 2 to clarify magnitudes of differences. The coefficients in the figure are exponentiated to increase interpretability: They may be interpreted as a ratio relative to the discrimination ratio of the reference category country of the United States. For instance, 1.26 indicates a discrimination ratio 26 percent higher than in the United States.

In models with basic controls (Table 3, model 2), France stands out with a discrimination ratio 43 percent higher than that of the United States (significantly different from the United States at $p < 0.001$). Sweden is next, with a discrimination ratio about 30 percent higher than that of the United States (significantly different from the United States at $p < 0.1$).¹⁰ Next highest are Canada, Great Britain, Belgium, the Netherlands, Norway, and the United States. Differences among these countries are not statistically significant. Finally, Germany shows lower levels of

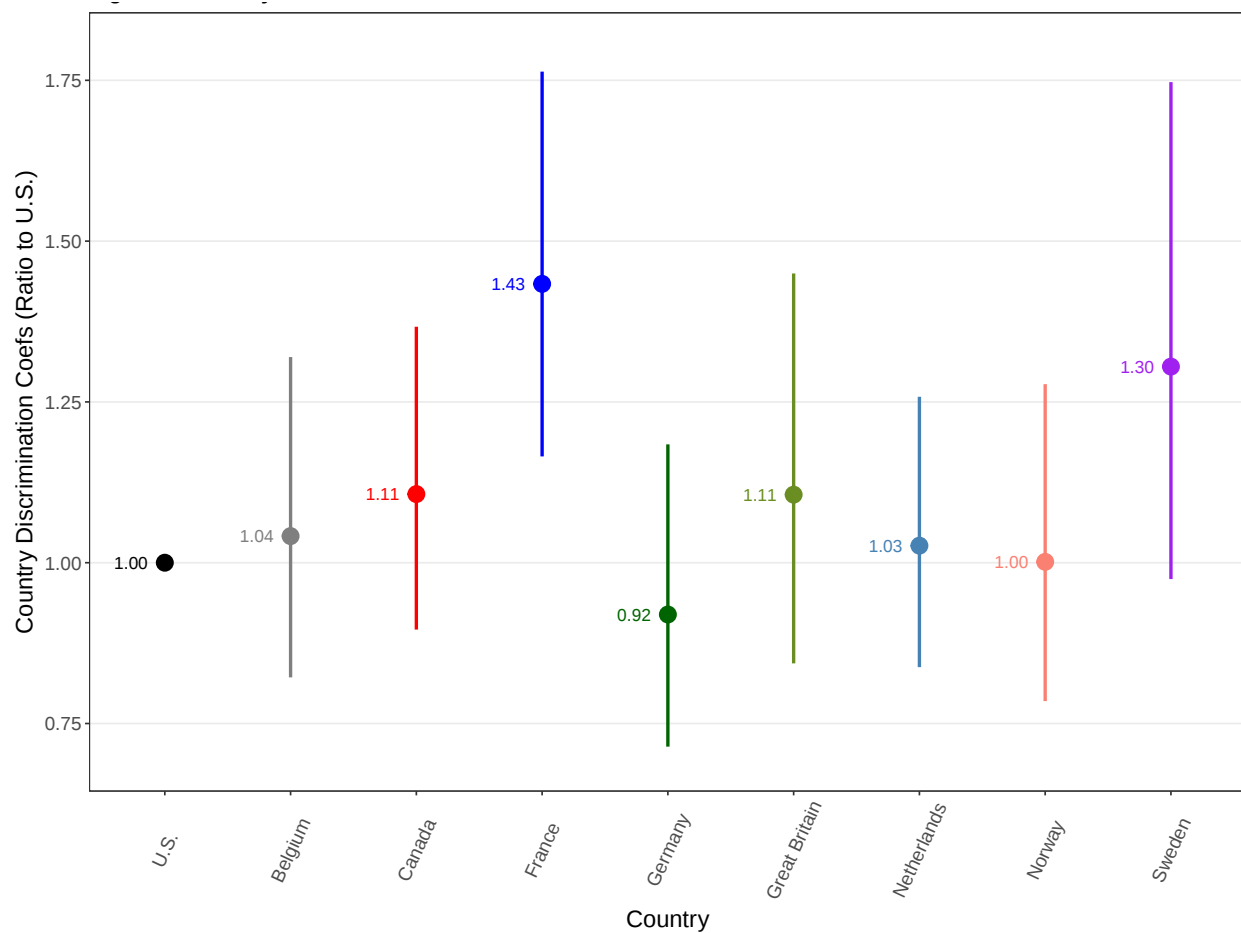


Figure 2: Country discrimination levels relative to the United States. Lines are 95 percent confidence intervals. Estimates are based on exponentiated coefficients from Table 3, model 2.

discrimination than the United States, with a discrimination ratio about 8 percent lower than that of the United States (not a statistically significant difference).

Minority group effects are presented in Figure 3 based on exponentiating the minority group coefficients in model 2 in Table 3. The coefficients in Figure 3 are relative to discrimination against persons who are African/black and can be interpreted as ratios of the discrimination ratio relative to African/black targets. The results show strong evidence of lower discrimination against European immigrants than persons who are African/black. By contrast, rates of discrimination seem rather similar in level among African/black, Middle Eastern/North African, and Asian minority groups. Discrimination against Latin American or Latino groups seems less than the other nonwhite groups but more than European groups, although this is not statistically significant from the others at $p < 0.05$.

Model 3 adds some controls for foreign characteristics, the applicants being foreign-born or having foreign credentials. None of these significantly predicts the outcome, which may in part reflect low variability on some measures (applicants in

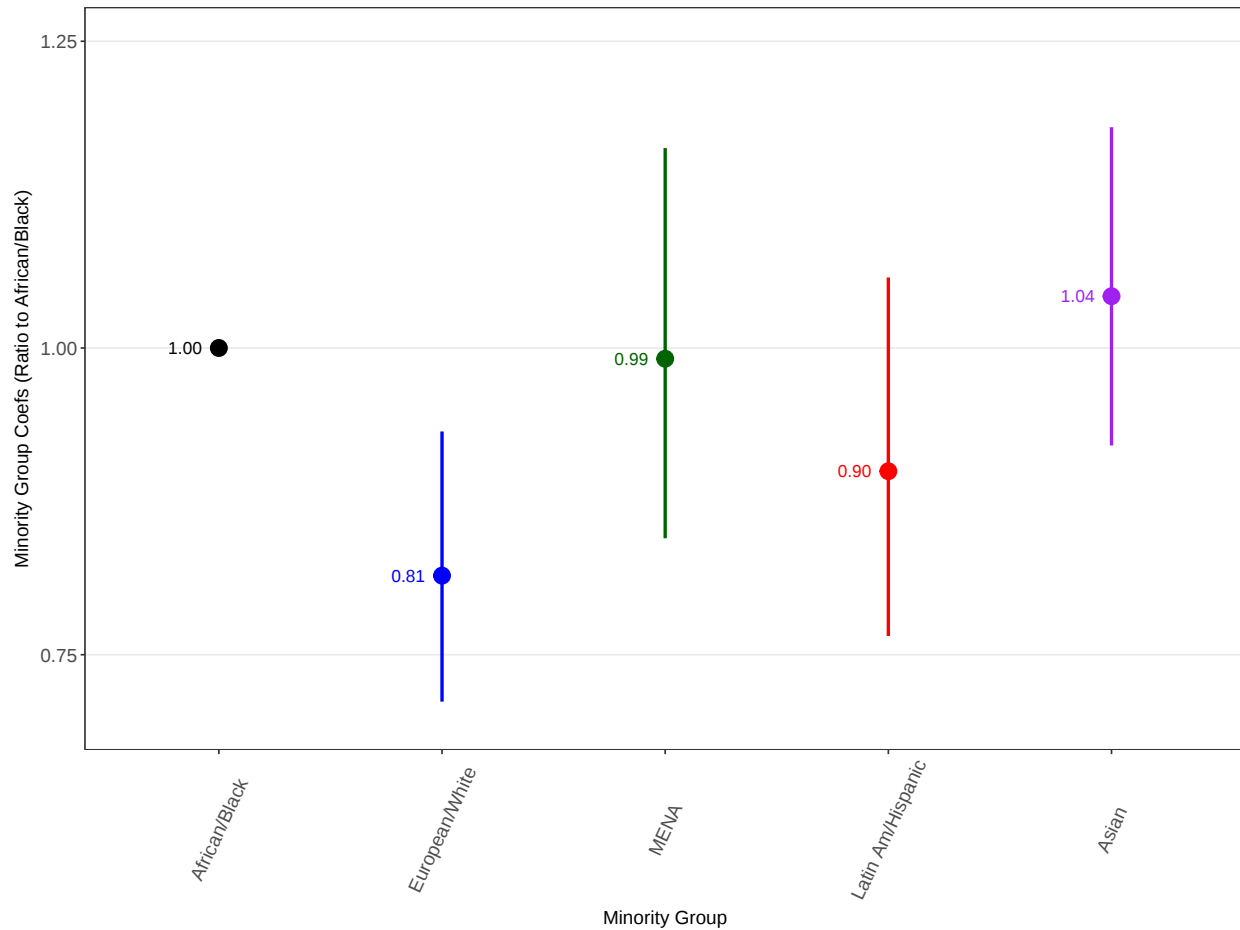


Figure 3: Minority group discrimination levels relative to Africans and/or blacks. Lines are 95 percent confidence intervals. Estimates are based on exponentiated coefficients from Table 3, model 2.

most studies are native born).¹¹ Country and group effect estimates are essentially unchanged by the controls.

Finally, model 4 adds two contextual characteristics as controls: local unemployment rates and immigrant share of the local population. Neither are statistically significant predictors, and both have small coefficients. This is contrary to some hypotheses regarding the importance of these characteristics in the group threat literature. We would have liked to have been able to include the share of specific racial-ethnic groups in the local area, but the lack of comparable cross-national data reporting on race and ethnicity made this impossible. Neither of these covariates explain any of the country or minority group differences.

On unemployment, we note the lack of an unemployment effect on discrimination is consistent with two prior studies based on field experiments that find no significant effect of unemployment levels on discrimination in hiring (Zschirnt and Ruedin 2016; Vuolo, Uggen and Lageson 2017). Minority hiring could still increase as unemployment declines because there are fewer white applicants with good

resumes applying for jobs (if few white applications with good qualifications are unemployed) rather than because the rate of discrimination changes.

Across all models, country difference point estimates tend to be larger than other predictors and change little as controls are introduced to the model. These differences are substantively large: The models imply that on average, white natives receive 75 percent to 102 percent more callbacks in France and Sweden than nonwhite minorities; in Germany, the United States, and Norway, they receive 22 to 41 percent more (model predictions are shown in online supplement Table S2). Country generally has larger effects than our other measured study-level characteristics.¹²

Analysis Checks

Interactions of Country, Minority Group, and Other Covariates

Our models in Table 3 assume that country and minority group effects are additive in predicting discrimination, not interactive. Table 4 contrasts fit statistics of different models that weaken this and other modeling assumptions. The first row of Table 4 shows fit statistics for the base model (Table 3, model 2). We can contrast this with the other models. The fit statistics shown are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), which evaluate model fit penalizing the complexity of the model (Raftery 1995). For both measures, smaller numbers indicate better fit.

The results indicate clearly better penalized fit without interactions of country and group. Levels of relative discrimination against minority groups tend to be similar within countries. Likewise, they provide evidence against several alternative specifications, such as allowing year-specific country trends or using dummy variables to represent decades instead of a linear year term in favor of our base specification of additive country and group effects.

Publication Bias

A potential problem in any meta-analysis is publication bias: that studies that do not find statistically significant differences might be less likely to be published and then less likely to be included in a meta-analysis (Sutton 2009). In Appendix 1 in the online supplement, we discuss steps we took to mitigate this potential problem and test for publication bias from the meta-analysis literature. We find statistically significant evidence of funnel-plot asymmetry that could reflect publication bias for some countries, but when we adjust for the bias, the changes in discrimination levels are very small and have no effect on substantive results.

Alterations to the Outcome Measure and Years Included

Sensitivity analyses showed that altering the years of data in our core sample did not alter our main results (see online supplement Table S1 for analysis using only studies since 1989). Using only the more recent years of data, our estimates are

Table 4: Fit statistics for models with interactions and alternative specifications.

	AIC	BIC	Model Parameters
Base Model, No Interactions (Table 3, Model 2)	62.8	132.7	24
Base Model and Interactions of Target Group by Country	88.0	192.1	37
Base Model with Dummies for Decade in Place of Linear Year	67.5	142.8	26
Base Model and Interactions of Year and Country	64.6	155.8	32
Base Model and Interactions of Decade and Country	76.1	180.1	37

Note: Lower numbers indicate better model fit.

basically similar. The one notable change is that Great Britain has a discrimination ratio similar to that of France and Sweden using only post-1989 studies. However, there are only three British studies since 1989, in contrast with seven before 1989, providing a thin basis for inference.

Discussion

In every country we consider, nonwhite applicants suffer significant disadvantage in receiving callbacks for interviews compared with white natives with similar job-relevant characteristics. This difference is driven by race, not immigrant status; our measures of native versus immigrant place of birth are not significant in predicting discrimination. White immigrants (and their descendants) are also disadvantaged relative to white natives but less so than nonwhites, and the difference between white immigrants and white natives is often small and statistically insignificant. We find fairly similar levels of discrimination against nonwhite groups irrespective of their specific origins: Persons of African descent, persons from the Middle East or North Africa, and persons of Asian descent (mostly South Asia in our data) experience roughly equal levels of discrimination. Broadly, our results are consistent with perspectives such as social dominance theory (Sidanius and Pratto 2001) that emphasize the pervasiveness of discrimination against nonwhites in Europe and North America. In these respects, we find a common pattern of discrimination across European and North American countries.

However, we find that the level of discrimination against minority groups varies considerably across countries. On average, whites receive 65 percent to 100 percent more callbacks in France and Sweden than nonwhite minorities; in Germany, the United States, and Norway, they receive 20 to 40 percent more. Differences by country are larger and more significant than most of the measured social and study factors we include. In the domain of hiring, some countries do discriminate more than others.

Not only are the differences in levels across countries large, but the ranking of countries defies most prior expectations. France is the country with the highest level of discrimination, followed by Sweden. By contrast, Germany, Norway, and the United States have lower rates of discrimination. The cross-national variation in hiring discrimination identified here does not correspond closely with other documented patterns of ethnic or racial inequality in the countries included. Our results are, for example, rather different from the ranking of countries according to

disparities in unemployment rates between immigrants and natives reported by the OECD (2015). The simple correlation between our national discrimination estimates and the ratio of unemployment of immigrants to natives gives a correlation of about 0.2.¹³ Of course, a country can have large-scale ethnoracial inequalities in employment outcomes while simultaneously having a low level of hiring discrimination. The gross inequalities reported by the OECD take no account, for example, of group differences in levels of human capital or effects of selective migration.

One might, in contrast, expect that our results would be more consistent with the cross-national variation in employment found in the ethnic penalties literature (e.g., Heath and Cheung 2007). Ethnic penalties are the gaps in employment or earnings between groups that remain after controlling for human capital and individual characteristics, such as gender and age. Indeed, estimates of national ethnic penalties in Western labor markets do correspond better with the discrimination estimates of our meta-analysis than do the unadjusted unemployment rates reported by OECD. Figure 4 graphs ethnic penalties by target group and country (from Heath and Cheung 2007) against our discrimination estimates (from Figure 1).¹⁴ The two measures are correlated at 0.50, suggesting a role of hiring discrimination in explaining racial and ethnic employment gaps.

To be sure, there are a number of additional mechanisms apart from discrimination that could generate ethnic penalties. For example, a lack of social capital and bridging ties to the majority group has been convincingly implicated in minority employment disadvantage (Lancee 2010; Zucotti and Platt 2017). Selectivity of migrants in labor market characteristics, grounded in national immigration laws and the nature of migration streams, is also likely to account for some group differences (Lee and Zhou 2015; van de Werfhorst and Heath 2019).

An important point to note about our analysis is that our estimates examine only hiring, not other aspects of the employment relationship. To the extent that processes of statistical discrimination or stereotyping are important contributors to hiring discrimination, the uncertainty inherent in hiring may make race and ethnicity a more important basis of discrimination at the point of hire than for other aspects of the employment relationship, such as wage setting and termination, for which employers have more complete information.

Likewise, the jobs included in field experiments of discrimination are typically those listed in public sources such as job databanks and newspapers advertisements. Our results do not well capture the situation of jobs advertised primarily through word of mouth or through less mainstream job sources such as ethnic newspapers. If members of a minority group in a country focus their job search on ethnic submarkets, they will likely experience less direct discrimination than we find. At the same time, studies show that ethnic-economy jobs tend to be on a secondary market, often with lower pay and benefits (Xie and Gough 2011). Access of ethnic minorities to broadly advertised positions is thus a critical component in the broader processes driving racial and ethnic inequality and immigrant incorporation.

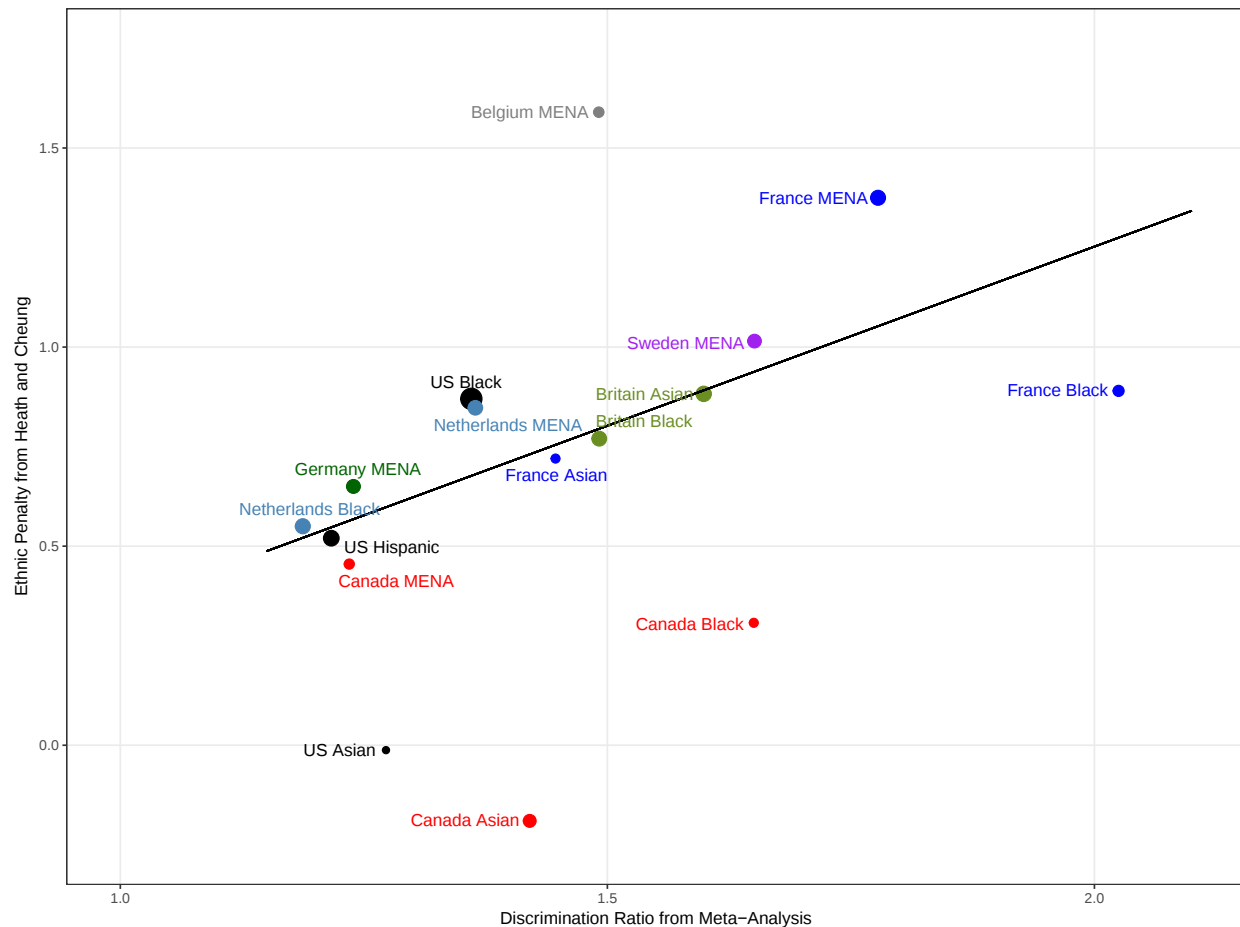


Figure 4: Ethnic penalty and discrimination by country and target group. The size of the symbol is proportional to meta-analysis weight. Country and minority group are indicated next to the symbol.

Implications

Although we are not able to definitively prove the reasons for the national variation we find, our results have implications for several theoretical arguments in the literature about factors driving discrimination.

Theories that emphasize persistent, direct effects of historical racial oppression that differ across countries—slavery and colonialism—may explain the common pattern of discrimination across groups but not the difference in levels across countries. That is, the ubiquity of discrimination against nonwhites (and low discrimination against white immigrants) may be the result of common cross-national histories linked to white supremacy. But national histories of slavery and colonialism are neither necessary nor sufficient conditions for a country to have relatively high levels of labor market discrimination. Some countries with colonial pasts demonstrate high rates of hiring discrimination, but several countries without extensive colonial pasts (outside Europe), such as Sweden, demonstrate similar levels. Likewise, the lower rates of discrimination against minorities in the United

States than we find for many European countries seem contrary to expectations that emphasize the primacy of connection to slavery in shaping the contemporary level of national discrimination. These results do not suggest that slavery and colonialism do not matter for levels of discrimination, rather they indicate that they matter in more complex ways than suggested by theories that posit simple, direct influences of the past on current discrimination.

High discrimination in the French labor market seems inconsistent with claims made by some scholars that discourse or measurement of race and ethnicity itself will tend to produce more discrimination by promoting “groupism” and group stereotypes (Sniderman and Hagendoorn 2007). The efforts in France not to measure or formally discuss race or ethnicity do not seem to have led to less discrimination. Some might claim that relatively high unemployment in the French labor market produces discrimination, but our analysis directly controls for local unemployment rates and finds this does not predict the discrimination ratio.

We note as well that the cross-national differences we find should not be read as primarily reflecting national levels of prejudice or as indicators of national levels of racism. Our discrimination measures are specific to hiring, and some evidence suggests national levels in discrimination in other outcomes may be different. For instance, we find low hiring discrimination in Germany,¹⁵ but Germany has not been found to be low on housing discrimination (Auspurg, Schneck, and Hinz 2018), suggesting weak antiminority prejudice may not account for this result. More likely, low discrimination in Germany could be a result of distinctive hiring practices in Germany: Employees typically submit far more extensive background information at initial application than in most other countries—including, for instance, high school transcripts and reports from apprenticeships (Weichselbaumer 2016). This may reduce the tendency of employers to assume lower skills and qualifications among nonwhite applicants, which is one potential source of discrimination. If so, this suggests the importance of high levels of individual information about applicants as a method to mitigate discrimination (c.f., Wozniac 2015; Auspurg et al. 2018).

Although we doubt that any one factor can account for the cross-national differences we find, we see considerable promise in the project of investigating the common underlying features that may be at work. To take one example, there is substantial cross-national variation in the institutional mechanisms through which employer discretion is constrained and the extent to which it is constrained at all. Few constraints are placed on employers’ ethnic consideration in hiring in France, which is largely due to the absence of monitoring or measurement along these lines. Likewise, in Sweden, reliance on employers and unions to oversee hiring practices (Törmkvist 2013), and a lack of monitoring of employee diversity among employers, may contribute to the relatively high levels of hiring discrimination. By contrast, the system in the United States provides for ethnic monitoring through affirmative action for federal contractors, legal penalties through the Equal Employment Opportunity Commission and private lawsuits, and institutional mechanisms that promote diversity through human resource practices at many large corporations. Although U.S. antidiscrimination laws include some ineffective and even counterproductive provisions (Berrey, Nelson, and Nielsen 2017), evidence also suggests

at least some parts of antidiscrimination laws and practices reduce discriminatory treatment (Leonard 1990; Kalev et al. 2006). The German system, with its detailed requirements for information about applicants' qualifications and its close articulation between education, training, and employment, follows a very different path for limiting the scope for employer discretion.

Future research is needed to convert these conjectures into firmly based explanations. For now, our contribution adds to the literature by demonstrating that patterns of discrimination vary sharply according to national context, by providing evidence against many commonly conceived explanations for this variation, and by suggesting some that may be important. Marshalling additional evidence to better understand the sources of country-level variation in hiring discrimination remains in the purview of future research.

Notes

- 1 At least colonialism outside of Europe, because Sweden for a time was an empire with territories in Northern Europe.
- 2 To be sure, some criticisms of field experiments have been raised by economists, such as Heckman and Siegelman (1992) and Neumark (2012). Perhaps the most telling criticism is that, in the case of in-person audit studies, the actors who implement the job application will not be blind to the nature of the experiment and may therefore act in ways that compromise validity. This concern is largely obviated by the large number of field experiments that rely on resumes only. Other criticisms that postulate statistical discrimination deriving from employers' beliefs about the means or variances of group differences in productivity have less force: From a legal point of view, statistical discrimination on the basis of group differences still counts as discrimination against the individual concerned.
- 3 The one field experimental study we know of conducted before 2016 that is cross-nationally comparative is that of Akintola (2010). Studies commissioned by the International Labour Organization used some elements of common design across countries but were conducted by different national teams at different points in time that often modified the base design quite significantly, making the results more similar to separate studies than a single cross-national study.
- 4 A problematic aspect of Zschirnt and Ruedin's (2016) analysis is that they treat subeffects from studies as though they are independent discrimination estimates. By doing this, they derive 624 effect sizes (discrimination estimates) from 42 correspondence studies, which they treat as 624 independent estimates of discrimination. These effect sizes are not independent for two reasons. First, in some cases, the effect sizes contrast different subsets of racial minority group members to the same majority control group, creating a strong form of dependence among the effect sizes because they are calculated partially from the same individuals (majority group members). Second, effect sizes from the same studies use many of the same procedures, such as having testers with similar education levels and aspects of similar resumes. In Hedges, Tipton, and Johnson's (2010) discussion of dependent effect sizes in meta-analysis, these correspond to "correlated" and "hierarchical" forms of dependence. The result of treating these effects as independent is that their standard errors are too small, and significance tests will have a high likelihood of producing type I errors.

- 5 Data, code, and a bibliographic list of studies included are available online at <https://sites.northwestern.edu/dmap/>.
- 6 In five studies, discrimination ratios were adjusted to account for a preapplication stage in which applicants called employers and asked if the job was still available. The adjustment is discussed in Appendix 2 in the online supplement.
- 7 For a clear introduction to meta-analysis and meta-regression, see Borenstein et al. (2009).
- 8 Focusing on the statistical significance of results for individual country and group comparisons would risk a high chance of type I errors absent in adjustment for multiple comparisons.
- 9 Coefficients can be interpreted as percentage differences by taking the coefficient and applying the formula $100*(\text{Exp}[B]-1)$.
- 10 This result is consistent with results of a correspondence study conducted by Akintola (2010) in Sweden and Canada using identical procedures that found Sweden had much higher discrimination in hiring than Canada. If we adjust p values for multiple comparisons using the Bonferroni method based on eight contrasts, France remains statistically significantly different from the United States, although the difference from Sweden is not statistically significant.
- 11 Results are unchanged if we enter these foreign characteristics one at a time in the model to avoid potential multicollinearity problems.
- 12 Fit statistics also confirm that country is more important than other sets of predictors. The r^2 of the base model (Table 3, model 2) is 36.2; dropping minority group dummies causes this to drop to 31.1, dropping country dummies (retaining minority group and controls) causes this to drop to 15.1, and dropping controls (retaining minority group and country) causes this to drop to 29.5. Dropping the country dummies causes by far the greatest decline in r^2 .
- 13 Country discrimination estimates are based on the coefficients of model 2 in Table 3.
- 14 Ethnic penalties are from Heath and Cheung's (2007) Tables 15.4 and 15.5. We averaged the ethnic penalties estimates over genders using target groups that matched the broad categories in each country's experimental discrimination studies. Ethnic penalty and discrimination estimates are available for the same groups for 15 country and group combinations ($n = 15$; Figure 4). Ethnic penalty estimates are not available for Norway. Estimates are weighted by the inverse variance of the discrimination ratio estimate from the meta-analysis. The ethnic penalties are based on logistic regression coefficients; to the extent that these may also capture unaccounted-for residual heterogeneity that differs by country, this should act as random measurement error.
- 15 The German point estimate is low, but Germany is not statistically significantly lower than the United States.

References

- Akintola, Abayomi Rasheed. 2010. "Hiring Discrimination in Racially Diverse Labour Markets: A Cross Country Study." Thesis, Karlstad Business School, Karlstad, University.
- Alba, Richard, and Nancy Foner. 2015. *Strangers No More: Immigration and the Challenges of Integration in North America and Western Europe*. Princeton, NJ: Princeton University Press. <https://doi.org/10.3917/popu.1503.0673>.
- American Anthropological Association. 1997. "Response to OMB Directive 15." https://s3.amazonaws.com/rdcms-aaa/files/production/public/FileDownloads/pdfs/cmtes/minority/upload/AAA_Response_OMB1997.pdf.

- American Sociological Association. 2002. "The Importance of Collecting Data and Doing Social Scientific Research on Race." American Sociological Association. http://www.asanet.org/images/press/docs/pdf/asa_race_statement.pdf.
- Auspurg, Katrin, Andreas Schneck, and Thomas Hinz. 2018. "Closed Doors Everywhere? A Meta-analysis of Field Experiments of Ethnic Discrimination in Rental Housing Markets." *Journal of Ethnic and Migration Studies* 45:95–114. <https://doi.org/10.1080/1369183x.2018.1489223>.
- Austin, Algernon. 2013. "The Unfinished March: An Overview." Economic Policy Institute. <http://www.epi.org/publication/unfinished-march-overview/>.
- Baert, Stein. 2018. "Hiring Discrimination, and Overview of (Almost) All Correspondence Experiments Since 2005." Pp. 63–77 in *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, edited by S. M. Gaddis. New York, NY: Springer International Publishing. https://doi.org/10.1007/978-3-319-71153-9_3.
- Beaman, Jean. 2017. *Citizen Outsider: Children of North African Immigrants in France*. Berkeley, CA: University of California Press. <https://doi.org/10.1525/luminos.39>.
- Berrey, Ellen, Robert L. Nelson, and Laura Beth Nielsen. 2017. *Rights on Trial: How Workplace Discrimination Law Perpetuates Inequality*. Chicago, IL: University of Chicago Press. <https://doi.org/10.1086/701591>.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *The American Economic Review* 94:991–1013. <https://doi.org/10.1257/0002828042002561>.
- Bertrand, Marianne, and Esther Duflo. 2016. "Field Experiments of Discrimination." Working Paper 22014, National Bureau of Economic Research.
- Blalock, Hubert. 1967. *Toward a Theory of Minority-Group Relations*. New York, NY: Wiley.
- Bonilla-Silva, Eduardo. 2006. *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality*. Lanham, MD: Rowman & Littlefield.
- Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-analysis*. 1st ed. Chichester, United Kingdom: Wiley. https://doi.org/10.1111/j.1751-5823.2009.00095_15.x.
- Branton, Regina, Erin C. Cassese, Bradford S. Jones, and Chad Westerland. 2011. "All Along the Watchtower: Acculturation Fear, Anti-Latino Affect, and Immigration." *The Journal of Politics* 73:664–79. <https://doi.org/10.1017/s0022381611000375>.
- Cooper, Harris, Larry V. Hedges, and Jeff C. Valentine. 2009. *The Handbook of Research Synthesis and Meta-Analysis*. New York, NY: Russell Sage Foundation.
- Dobbin, Frank. 2011. *Inventing Equal Opportunity*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1017/s0022050712000472>.
- European Union Agency for Fundamental Rights. 2008. *Annual Report 2008*. Vienna, Switzerland: European Union Agency for Fundamental Rights. <https://doi.org/10.5040/9781472563231.ch-010>.
- Feagin, Joe R., and Melvin P. Sikes. 1994. *Living with Racism: The Black Middle-Class Experience*. Boston, MA: Beacon Press.
- Fisher, Zachary, and Elizabeth Tipton. 2015. "Robumeta: An R-package for Robust Variance Estimation in Meta-analysis." Cornell University. <https://arxiv.org/abs/1503.02220>.
- Fredrickson, George M. 2002. *Racism: A Short History*. Princeton, NJ: Princeton University Press.

- Gaddis, Michael. 2017. "How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies." *Sociological Science* 4:469–89. <https://doi.org/10.15195/v4.a19>.
- Golder, Matt. 2016. "Far Right Parties in Europe." *Annual Review of Political Science* 19:477–97. <https://doi.org/10.1146/annurev-polisci-042814-012441>.
- Gould, Stephen Jay. 1996. *The Mismeasure of Man*. New York, NY: W. W. Norton.
- Heath, Anthony, Thomas Liebig, and Patrick Simon. 2013. "Discrimination against Immigrants – Measurement, Incidence and Policy Instruments." OECD iLibrary. http://www.oecd-ilibrary.org/social-issues-migration-health/international-migration-outlook-2013_migr_outlook-2013-en.
- Heath, Anthony F., and Sin Yi Cheung. 2007. *Unequal Chances: Ethnic Minorities in Western Labour Markets*. Oxford, United Kingdom: Oxford University Press for the British Academy. <https://doi.org/10.5871/bacad/9780197263860.001.0001>.
- Heckman, James J., and Peter Siegelman. 1992. "The Urban Institute Audit Studies: Their Methods and Findings," Pp. 187–258 in *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by M. Fix and R. J. Struyk. Lanham, MD: Urban Institute Press.
- Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12:101–16.
- Hedges, Larry V., Elizabeth Tipton, and Matthew C. Johnson. 2010. "Robust Variance Estimation in Meta-regression with Dependent Effect Size Estimates." *Research Synthesis Methods* 1:39–65. <https://doi.org/10.1002/jrsm.5>.
- Jonsson Jan O., Frank Kalter, and Frank van Tubergen. 2018. "Studying Integration: Ethnic Minority and Majority Youth in Comparative Perspective." Pp.3–39 in *Growing up in Diverse Societies: The Integration of the Children of Immigrants in England, Germany, the Netherlands, and Sweden*, edited by F. Kalter, J. O. Jonsson, F. van Tubergen, and A. Francis. Oxford, United Kingdom: Oxford University Press. <https://doi.org/10.5871/bacad/9780197266373.003.0001>.
- Kalev, Alexandra, Frank Dobbin, and Erin Kelly 2006. "Best Practice or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies." *American Sociological Review* 71:589–617. <https://doi.org/10.1177/000312240607100404>.
- Lancee, Bram. 2010. "The Economic Returns of Immigrants' Bonding and Bridging Social Capital: The Case of the Netherlands." *International Migration Review* 44:202–26. <https://doi.org/10.1111/j.1747-7379.2009.00803.x>.
- Lee, Jennifer, and Min Zhou. 2015. *The Asian American Achievement Paradox*. New York, NY: Russell Sage Foundation.
- Legewie, Joscha. 2013. "Terrorist Events and Attitudes toward Immigrants: A Natural Experiment." *American Journal of Sociology* 118:1199–245. <https://doi.org/10.1086/669605>.
- Leonard, Jonathan S. 1990. "The Impact of Affirmative Action Regulation and Equal Employment Law on Black Employment." *Journal of Economic Perspectives* 4:439–463. <https://doi.org/10.1257/jep.4.4.47>.
- Meyer, John W., John Boli, George M. Thomas, and Francisco O. Ramirez. 1997. "World Society and the Nation-State." *American Journal of Sociology* 103:144–81. <https://doi.org/10.1086/231174>.
- National Research Council. 2004. *Measuring Racial Discrimination*. Washington, DC: National Academies Press.

- Neumark, David. 2012. "Detecting Discrimination in Audit and Correspondence Studies." *Journal of Human Resources* 47:1128–57. <https://doi.org/10.3368/jhr.47.4.1128>.
- Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature* 56:799–866. <https://doi.org/10.1257/jel.20161309>.
- O'Connell, Heather A. 2012. "The Impact of Slavery on Racial Inequality in Poverty in the Contemporary U.S. South." *Social Forces* 90:713–34. <https://doi.org/10.1093/sf/sor021>.
- Organisation for Economic Co-operation and Development. 2015. *Indicators of Immigrant Integration 2015: Settling In*. Paris, France: OECD Publishing. <http://dx.doi.org/10.1787/9789264234024-en>.
- Pager, Devah, and Lincoln Quillian. 2005. "Walking the Talk? What Employers Say Versus What They Do." *American Sociological Review* 70:355–80. <https://doi.org/10.1177/000312240507000301>.
- Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *Annals of the American Academy of Political and Social Science* 609:104–33. <https://doi.org/10.1177/0002716206294796>.
- Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34:181–209. <https://doi.org/10.1146/annurev.soc.33.040406.131740>.
- Pager, Devah, Bart Bonikowski, and Bruce Western. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74:777–99. <https://doi.org/10.1177/000312240907400505>.
- Quillian, Lincoln. 1995. "Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-Immigrant and Racial Prejudice in Europe." *American Sociological Review* 60:586–611. <https://doi.org/10.2307/2096296>.
- Quillian, Lincoln. 2006. "New Approaches to Understanding Racial Prejudice and Discrimination." *Annual Review of Sociology* 32:299–328. <https://doi.org/10.1146/annurev.soc.32.061604.123132>.
- Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn Midtbøen. 2017. "Meta-analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring over Time." *The Proceedings of the National Academy of Sciences of the United States of America* 114:10870–5. <https://doi.org/10.1073/pnas.1706255114>.
- Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111–63. <https://doi.org/10.2307/271063>.
- Raudenbush, Stephen. 2009. "Analyzing Effects Sizes: Random Effects Coding." in *The Handbook of Research Synthesis and Meta-Analysis*, Cooper, edited by H. M. Cooper, L. V. Hedges, and J. C. Valentine. New York, NY: Russell Sage Foundation.
- Riach, Peter A., and Judith Rich. 2002. "Field Experiments of Discrimination in the Market Place." *The Economic Journal* 112:F480–518. <https://doi.org/10.1111/1468-0297.00080>.
- Schlueter, Elmar, and Peer Scheepers. 2010. "The Relationship between Outgroup Size and Anti-Outgroup Attitudes: A Theoretical Synthesis and Empirical Test of Group Threat and Intergroup Contact Theory." *Social Science Research* 39:285–95. <https://doi.org/10.1016/j.ssresearch.2009.07.006>.
- Semyonov, Moshe, Rebeca Raijman, and Anastasia Gorodzeisky. 2006. "The Rise of Anti-Foreigner Sentiment in European Societies, 1988–2000." *American Sociological Review* 71:426–49. <https://doi.org/10.1177/000312240607100304>.

- Sidanius, Jim, and Felicia Pratto. 2001. *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge, United Kingdom: Cambridge University Press. <https://doi.org/10.1017/cbo9781139175043.014>.
- Simon, Patrick. 2008. "The Choice of Ignorance: The Debate on Ethnic and Racial Statistics in France." *French Politics, Culture & Society* 26-1:7–31. <https://doi.org/10.3167/fpcs.2008.260102>.
- Sniderman, Paul, and Louk Hagendoorn. 2007. *When Ways of Life Collide: Multiculturalism and Its Discontents in the Netherlands*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1007/s11562-008-0039-x>.
- Sutton, Alex J. 2009. "Publication bias." Pp. 435–52 in *The Handbook of Research Synthesis and Meta-Analysis*, edited by H. M. Cooper, L. V. Hedges, and J. C. Valentine. New York, NY: Russell Sage Foundation.
- Taylor, Marylee. 1998. "How White Attitudes Vary with the Racial Composition of Local Populations: Numbers Count." *American Sociological Review* 63:512–35. <https://doi.org/10.2307/2657265>.
- Törnkvist, Ann. 2013. "Is the Swedish Model the Fairest of them All?" *The Local*. <https://www.thelocal.se/20131112/is-the-swedish-model-the-fairest-of-them-all-labour-employment>.
- van de Werfhorst, Herman G., and Anthony Heath. 2019. "Selectivity of Migration and the Educational Disadvantages of Second-Generation Immigrants in Ten Host Societies." *European Journal of Population* 35:347–78. <https://doi.org/10.1007/s10680-018-9484-2>.
- Viechtbauer, Wolfgang. 2010. "Conducting Meta-analyses in R with the Metafor Package." *Journal of Statistical Software* 36:1–48.
- Vuolo, Michael, Christopher Uggen, and Sarah Lageson. 2017. "Race, Recession, and Social Closure in the Low-Wage Labor Market: Experimental and Observational Evidence." *Research in the Sociology of Work* 30:141–83. <https://doi.org/10.1108/s0277-283320170000030007>.
- Weichselbaumer, Doris. 2016. "Discrimination against Female Migrants Wearing Headscarves." Discussion Paper 10217, IZA.
- Wilson, William Julius. 2012. *The Declining Significance of Race: Blacks and Changing American Institutions*. 3rd ed. Chicago, IL: University Of Chicago Press. <https://doi.org/10.1093/sw/24.4.352-a>.
- Winant, Howard. 2001. *The World Is a Ghetto: Race and Democracy Since World War II*. New York, NY: Basic Books.
- Wozniac, Abigail. 2015. "Discrimination and the Effects of Drug Testing on Black Employment." *The Review of Economics and Statistics* 97:548–66.
- Xie, Yu, and Margaret Gough. 2011. "Ethnic Enclaves and the Earnings of Immigrants." *Demography* 48:1293–315. <https://doi.org/10.1007/s13524-011-0058-8>.
- Zhou, Guang Yong. 2007. "One Relative Risk versus Two Odds Ratios: Implications for Meta-analyses Involving Paired and Unpaired Binary Data." *Clinical Trials* 4:25–31. <https://doi.org/10.1177/1740774506075667>.
- Zschirnt, Eva, and Didier Ruedin. 2016. "Ethnic Discrimination in Hiring Decisions: A Meta-analysis of Correspondence Tests 1990–2015." *Journal of Ethnic and Migration Studies* 42:1115–34. <http://dx.doi.org/10.1080/1369183X.2015.1133279>.
- Zucotti, Carolina V., and Lucinda Platt. 2017. "Does Neighbourhood Ethnic Concentration in Early Life Affect Subsequent Labour Market Outcomes? A Study across Ethnic Groups

in England and Wales." *Population, Space and Place* 23:e2041. <https://doi.org/10.1002/psp.2041>.

Acknowledgments: We have received financial support for this project from the Russell Sage Foundation and the Institute for Policy Research at Northwestern University. We thank Larry Hedges for methodological advice. We dedicate this article to Devah Pager, who learned a little from us and taught us much more.

Lincoln Quillian: Department of Sociology, Northwestern University.
E-mail: l-quillian@northwestern.edu.

Anthony Heath: Centre for Social Investigation, Nuffield College.
E-mail: anthony.heath@nuffield.ox.ac.uk.

Devah Pager: Deceased, formerly Department of Sociology, Harvard University.

Arnfinn H. Midtbøen: Institute for Social Research, Oslo, Norway.
E-mail: a.h.midtboen@samfunnsforskning.no.

Fenella Fleischmann: Interdisciplinary Social Science, Utrecht University.
E-mail: F.Fleischmann@uu.nl.

Ole Hexel: Department of Sociology, Northwestern University, and Observatoire Sociologique du Changement, Sciences Po, Paris, France.
E-mail: ole.hexel@u.northwestern.edu.