

Koopmans, Ruud; Veit, Susanne; Yemane, Ruta

Article — Published Version

## Taste or statistics? A correspondence study of ethnic, racial and religious labour market discrimination in Germany

Ethnic and Racial Studies

**Provided in Cooperation with:**

WZB Berlin Social Science Center

*Suggested Citation:* Koopmans, Ruud; Veit, Susanne; Yemane, Ruta (2019) : Taste or statistics? A correspondence study of ethnic, racial and religious labour market discrimination in Germany, Ethnic and Racial Studies, ISSN 1466-4356, Taylor & Francis, London, Vol. 42, Iss. 16, pp. 233-252, <https://doi.org/10.1080/01419870.2019.1654114>

This Version is available at:

<https://hdl.handle.net/10419/205261>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS



# Taste or statistics? A correspondence study of ethnic, racial and religious labour market discrimination in Germany

Ruud Koopmans, Susanne Veit and Ruta Yemane

WZB Berlin Social Science Center, Berlin, Germany

## ABSTRACT

In this study we compare rates of discrimination across German-born applicants from thirty-five ethnic groups in which various racial and religious treatment groups are embedded, this study allows us to better distinguish taste and statistical sources of discrimination, and to assess the relative importance of ethnicity, phenotype and religious affiliation as signals triggering discrimination. The study is based on applications to almost 6,000 job vacancies with male and female applicants in eight occupations across Germany. We test taste discrimination based on cultural value distance between groups against statistical discrimination based on average education levels and find that discrimination is mostly driven by the former. Based on this pattern, ethnic, racial and religious groups whose average values are relatively distant from the German average face the strongest discrimination. By contrast, employers do not treat minority groups with value patterns closer to Germany's different from ethnic German applicants without a migration background.

**ARTICLE HISTORY** Received 8 February 2019; Accepted 1 August 2019

**KEYWORDS** Ethnic discrimination; taste-based and statistical discrimination; field experiment; race and religion; value distance; education levels

## Introduction

Large numbers of correspondence studies have investigated labour-market discrimination against ethnic, religious and racial minorities (see e.g. the reviews by Bertrand and Duflo 2017; Rich 2014; Zschirnt and Ruedin 2016). In such studies, researchers send out comparable applications of fictitious candidates to real job openings but vary the characteristics of interest (e.g. gender and ethnicity). Different response rates (commonly referred to as “call-back”) provide causal evidence of discrimination (for overviews, see Gaddis 2018; Neumark 2016; Pager 2007). Across a wide range of countries, minority

**CONTACT** Ruta Yemane  [ruta.yemane@wzb.eu](mailto:ruta.yemane@wzb.eu)

 Supplemental data for this article can be accessed at <https://doi.org/doi:10.1080/01419870.2019.1654114>

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

groups and labour-market segments, these studies have demonstrated that candidates with the same observable productivity-relevant characteristics are nonetheless treated differently, thus proving the existence of labour-market discrimination beyond a reasonable doubt.

Correspondence studies have been much less successful, however, in isolating the reasons behind discrimination. Studies in Germany (Kaas and Manger 2012; Schneider, Yemane, and Weinmann 2014) have for instance shown that Turkish ethnics are discriminated, but it remains unclear whether this discrimination is based on assumptions about the average group productivity of Turkish ethnics or on anti-Turkish preferences void of any productivity-relevant empirical basis. Moreover, what triggers anti-Turkish biases? Is it a general anti-immigrant or pro-German bias that Turkish applicants suffer as much as other immigrant groups? Or is it some specific bias? And if so, is this related to their presumed Muslim religion or to phenotypical differences?

Answering such questions is crucial because discrimination rates vary strongly across ethnic (Oreopoulos 2011), racial (Quillian et al. 2017), phenotypical (Saeed, Maqsood, and Rafique 2019) and religious groups (Wright et al. 2013). In this paper, we use an innovative research design that includes thirty-five ethnic groups and different racial phenotypes and religious affiliations to analyse patterns of hiring discrimination in Germany. We test hypotheses about statistical and taste discrimination by analysing to what extent discrimination patterns vary with education levels or with cultural value differences across ethnic, racial and religious groups.

### **Taste-based and statistical discrimination**

The literature distinguishes two fundamentally different sources of discrimination: preferences (taste-based discrimination) and productivity concerns (statistical discrimination) (e.g. Guryan and Charles 2013 or Neumark 2016). Taste-based discrimination (Becker 1957) refers to bias that is unrelated to productivity concerns. If employers prefer working with majority members, they will invite them more often to job interviews than minorities. Employees or customers may have preferences for working with or being served by members of the majority group even when the employer has no such preference herself.

Statistical discrimination (Aigner and Cain 1977; Arrow 1973; Phelps 1972), by contrast, is related to productivity concerns. Information contained in job application materials is an imperfect signal of an applicant's true productivity and therefore employers, who have no personal experience with candidates, may additionally rely on signals of group membership that are not intrinsically causally related to unobserved productivity but that are empirically correlated with it. Such signals may include age and gender, but also race, ethnicity, or

religion. If an employer must choose between two candidates with equivalent observable skills, he may choose the candidate belonging to a group with higher average productivity on the assumption that the unobserved productivity component is likely to be higher for that candidate.

The distinction between taste and statistical discrimination is not just of theoretical interest, but also for designing adequate policy responses. If discrimination is taste-based, efficient measures would need to reduce the anti-minority bias, e.g. by way of anti-racist campaigns or diversity trainings. By contrast, if discrimination is statistical, discrimination is best combated by measures aiming at the empirical grounds on which it is based, e.g. by removing barriers and providing support programmes to increase minorities' qualifications.

Past studies have tested the hypothesis drawn from theories of statistical discrimination that providing more information about the individual productivity of candidates should reduce the callback gap between ethnic or racial groups (e.g. Agerström et al. 2012; Baert and De Pauw 2014; Bartoš et al. 2016; Edo, Jacquemet, and Yannelis 2013). Most studies indeed find less discrimination when richer information about individual productivity is provided, but generally, some discrimination remains even under the high-information treatment. The absence of empirical evidence in support of statistical discrimination or the remaining level of unexplained discrimination is sometimes interpreted as support for taste-based discrimination. But, to the best of our knowledge, there is no previous field experiment that directly tests taste-based discrimination.

At their core, however, theories of statistical and taste discrimination are not about varying levels of individual information, but about varying assumptions or preferences employers have about the groups applicants belong to. The design of most past correspondence studies does not allow an analysis of variation on the group level. Studies typically send two applications to the same vacancy, one by a majority applicant and one by a minority applicant, who in all other respects are equivalent in terms of observable skills. For instance, many US studies contrast black and white candidates (e.g. Bertrand and Mullainathan 2004, Gaddis 2014; Jacquemet and Yannelis 2012), whereas many European studies have contrasted Arab (e.g. Blommaert, Coenders, and Van Tubergen 2014; Derous, Ryan, and Nguyen 2012) or Turkish (e.g. Baert et al. 2015; Kaas and Manger 2012) applicants to majority applicants. This design is perfectly suited to demonstrate the occurrence of discrimination, but not to determine its causes or to generalize beyond the racial or ethnic groups studied. Statistical and taste discrimination implies competing hypotheses about group characteristics, which can never be answered by dichotomous comparisons of just one minority group to the majority. However, most designs compare only one minority group to a majority group. Some compare three (e.g. Pager 2003) or up to five (e.g. Booth,

Leigh, and Varganova 2012; Oreopoulos 2011) groups, but that gives only marginally more leverage to test competing hypotheses.

Employers who discriminate draw on signals of group membership to derive stereotypical assumptions about average characteristics of groups. These assumptions may have no basis in empirical facts at all or they may refer to empirically existing differences across groups. In relation to statistical discrimination, empirically unfounded assumptions about differences in group productivity have been referred to as “error discrimination” (England and Lewin 1989). Unlike forms of statistical discrimination based on at least partly valid assumptions about group differences, error discrimination lowers the efficiency of hiring decisions because it leads employers to devalue suitable candidates based on mistaken assumptions about the groups they belong to. We are not aware of a similar distinction between different forms of taste discrimination, probably because taste discrimination is regarded as a form of irrational behaviour per se. From an economic perspective, taste discrimination is indeed by definition irrational because it brings concerns unrelated to productivity into the hiring decision. But from other than economic perspectives there may be a rationale behind taste discrimination. Sociological studies have revealed strong preferences to interact with people who are socially and culturally similar (“homophily”, see McPherson, Smith-Lovin, and Cook 2001) – which may extend to the realm of business and work, since similarity is expected to facilitate communication and coordination by increasing the likelihood of similar experiences, goals and preferences (Byrne 1997; Montoya, Horton, and Kirchner 2008).

## Hypotheses

In this paper, we consider three signals of group membership: *ethnicity*, *race* and *religious affiliation*. To derive testable and empirically distinguishable hypotheses about statistical and taste discrimination, we assume that discrimination is at least in part based on existing differences across groups that can be measured and of which employers have some knowledge. Then we can predict that under statistical discrimination invitation rates should vary as a function of empirically observable average levels of productivity-related measures across groups. Similarly, under taste discrimination invitation rates should vary with empirically observable differences across groups regarding traits related to social and cultural preferences and tastes. These considerations lead to the following two hypotheses:

H1 (*statistical discrimination*): Discrimination is a function of group averages in productivity-relevant characteristics. The lower the average productivity of a group, the higher will be the rate of hiring discrimination against applicants belonging to that group.

H2 (*taste discrimination*): The greater the difference between the cultural value patterns of a group and those of the majority group, the higher will be the rate of hiring discrimination against applicants belonging to that group.

## Data and method

### Sample

Our sample consists of 5,819 applications to an equal number of vacancies advertised across Germany between October 2014 and April 2016 (for more details see the technical report: Veit and Yemane 2018). We used an unpaired design in order to accommodate many different treatments (e.g. gender, ethnicity, religion and phenotype). The sample includes male as well as female applicants who were born in Germany, had the German nationality, and had received their entire education in Germany. Vacancies were drawn from the website of the Federal German Employment Agency.

We applied to positions in eight medium-skilled occupations that require formal training. Because the labour market is partly segregated by gender, four of these were mixed-gender occupations, with roughly similar numbers of males and females (hotel receptionist, cook, salesperson and industrial office clerk); two were strongly male-dominated occupations (mechatronics fitter and plant mechanic for sanitary, heating and air conditioning systems); and two were strongly female-dominated (medical assistant and dental assistant). To avoid biases due to the atypicality of the gender of applicants for occupations that are strongly gender-skewed, we only sent male, respectively female applications to the four occupations with strong gender biases (see Online Supplement Table S1).

As is the norm in Germany, our applications not only included a motivation letter and a CV, but also full copies of vocational training certificates and secondary school diplomas. Further, it is customary to include letters of reference from previous employers and, importantly, a photograph of the applicant. The photo requirement makes it possible to signal racial phenotype directly rather than using the indirect and imprecise signal of typical minority names (see Gaddis 2017).

To draw conclusions on the plausible causes of group differences we consider three potentially relevant signals of group membership: *ethnicity*, *race* and *religious affiliation*. Importantly, since we let these different signals vary independently wherever this is empirically plausible, we can tease apart the separate effects of ethnic, racial and religious signals, which are normally conflated in correspondence studies. Moreover, because we substantially increased the number of ethnic groups in comparison to previous correspondence studies, our design allows us to apply a multivariate regression approach to test hypotheses drawn from theories of statistical and taste discrimination.

## ***Treatments, covariates and dependent variable***

### ***Ethnicity***

We selected ethnic groups from thirty-four countries of origin as well as a German comparison group. German and Turkish ethnics each make up a quarter of the sample, and the remainder is divided among the other thirty-three groups (see Online Supplement Table S2). Turkish ethnics were over-sampled to allow comparisons with results of earlier correspondence studies in Germany (e.g. Kaas and Manger 2012 or Schneider, Yemane, and Weinmann 2014). Except for the Turkish ethnics, we are not interested in deriving point estimates for single ethnic groups. The chosen ethnicities include the largest minority groups in Germany (e.g. Turkish, Bosnian, Polish, Russian and Italian ethnics) as well as several small and medium-sized ethnic groups that were selected to obtain enough variation across variables of interest. We signalled ethnicity in the application documents by names “typical” for the respective ethnic group, as well as, in the skills section of the CV, by indicating (in addition to German) a second mother tongue, e.g. “Luganda (Uganda, mother tongue)” or “English (USA, mother tongue)” (see for more details Veit and Yemane 2018).

### ***Racial phenotype***

Racial distinctions are arbitrary and boundaries fuzzy. Nevertheless, people racially categorize others and these categories may form a basis for discrimination. We use the three most widely recognized categories: white, black and Asian. These categories themselves, of course, contain phenotypical diversity; we, therefore, captured racial phenotypes by more than one photo. Altogether, twenty-eight photos were used, fourteen for men and women, respectively. For the white phenotype, we use six photos for each gender (varying by hair colour and skin tone), for the black phenotype four photos (varying by skin tone and hair texture), and for the Asian phenotype also four photos (varying by skin tone and facial features). We conducted two pre-tests, one for attractiveness and one on the plausibility of photos referring to members of different ethnic groups (see Veit and Yemane 2018). Based on the attractiveness pre-test two photos were replaced by alternative ones. Based on our plausibility pre-test we excluded implausible combinations of photos and ethnic groups. For instance, we include black Egyptians and Moroccans, but not black Turks. Online Supplement Table S2 shows which phenotypical treatments were used for each ethnic group. In addition, Table S4 provides the results of a post-test for the attractiveness of the final selection of photos.

### ***Religion***

Following other studies (e.g. Wallace, Wright, and Hyde 2014) we signalled religious affiliation by engagement in a voluntary association. For half of

the sample this engagement was non-religious (referring to a not further specified social association: "Sozialverein Aktiv e.V.") and for the other half, we indicated a religious affiliation (Christian, Muslim, Buddhist or Hindu: e.g. "Christlicher Sozialverein Aktiv e.V."). Where possible, we use more than one religious affiliation within the same ethnic group – using as a cut-off criterion that at least five per cent of the population of the country of origin must belong to religion for it to be included as a treatment. We thus replicate on a wider scale the innovation of using multiple religious backgrounds within one ethnic group, pioneered by Adida, Laitin, and Valfort (2016), Drydakis (2010), Pierné (2013) and Wright et al. (2013). Online Supplement Table S2 shows which religion treatments were used for each ethnic group. Because Hindu and Buddhist treatments are only plausible for a small number of ethnic groups (Hindu for Indians and Trinidadians and Buddhist for Japanese, South Koreans, Chinese, Vietnamese and Malaysians), we take them together in the empirical analyses.

### ***Other treatments***

Two other treatments that were randomly assigned will be used as control variables: the inclusion of a reference letter and school grades (see Veit and Yemane 2018). Half of the sample consists of applications with good grades whereas the other half has satisfactory grades.

## ***Group-level covariates: education levels and value distance***

### ***Level of education***

In order to test our hypothesis on statistical discrimination, we relate callback to education as an indicator of average productivity, using data from the 2012 Mikrozensus<sup>1</sup> (a one per cent representative sample of the German population). Average levels of education are calculated on a four-point scale (no education, only primary education, secondary and tertiary education).<sup>2</sup>

### ***Emancipative values***

As an indicator of taste discrimination, we use cultural distance in terms of emancipative values. We draw on the latest available data on average values in countries of origin, assuming that differences in values across immigrant groups will mirror to some extent those among their countries of origin.<sup>3</sup> We use the emancipative value dimension distinguished in the World Values Survey and the European Values Survey (see EVS 2015; WVS 2015, and the online appendix of Welzel 2013).<sup>4</sup> Instead of the value score itself, we use the absolute distance to the German score to obtain the value distance from Germany. Emancipative (vs. obedient) values measure four domains of emancipative orientations, ranging from support for freedom of speech and popular influence in national, local and job affairs (voice),

tolerance towards divorce, abortion and homosexuality (choice), personal autonomy as desired qualities in child-rearing (autonomy), to gender equality values (equality; e.g. "On the whole, men make better political leaders than women do").

For each covariate, we assume that employers' assumptions about group traits are triggered by the combined signals of ethnicity, race and religion. For instance, we suppose that employers estimate white Christian Egyptians and black Muslim Egyptians to have certain commonalities because of their Egyptian ethnic origin, but also to differ because of their different phenotypes and religions. Ideally, we would, therefore, need statistical data on productivity and values for each combination of signals, but the German Mikrozensus and the World and European Value Surveys do not provide this level of detail.

We, therefore, calculated scores for Christians, Muslims and Hindus or Buddhists by averaging over all countries represented in the Mikrozensus and the World and European Values Surveys with respectively Christianity, Islam and Hinduism or Buddhism as the dominant religion. In order to arrive at representative scores for the German context, we first defined for all origin groups in the Mikrozensus their dominant religion<sup>5</sup> and used this data to calculate mean education scores and emancipative values for each religious group. Thus, Turkish scores weigh more heavily in the calculation of the Muslim average than, e.g. Pakistani scores, because there are many times more people of Turkish than of Pakistani origin in Germany (and thus in the Mikrozensus) and concomitantly employer stereotypes of "Muslims" in Germany are more strongly driven by perceptions of Turks than of Pakistani.

We then similarly calculated averages for whites, blacks and Asians across all nationality groups where the respective race predominates.<sup>6</sup> The final measures used in the analyses represent each intersection of ethnicity, phenotype and religion separately, by taking the average of the ethnicity, race and religion scores to which an applicant belongs. White Christian Egyptians, therefore, get a different score (based on the average of the white, Christian and Egyptian scores) than black Muslim Egyptians (based on the average of the black, Muslim and Egyptian scores). In the no religious signal treatment, averages are calculated using the score for the dominant religion of the country of origin. This is justified because preliminary analyses showed that, for each of the three religious' groups, there were no significant callback differences between the no religion treatment and the majority religion treatment (e.g. Egyptians with no religion treatment vs. Egyptians with a Muslim treatment). Significant callback differences only resulted when an applicant had a minority religious signal (e.g. Muslim Bulgarians).

To demonstrate that our results also hold when we use more straightforward measures of education levels and value distance that are directly taken from the Mikrozensus and the European and World Values Surveys,

we also show results for simplified measures that vary only across the 35 ethnic groups (rather than the 132 intersections of ethnicity, race and religion). Online Supplement Table S3 provides the mean values of all group-level covariates across all groups.

### ***Dependent variable: positive callback***

Our dependent variable is a response from employers that indicates an interest in the candidate, which we refer to as “positive callback”. Most of these were straightforward invitations to a job interview, but the category also includes responses by email or phone that requested a return call or further information by email. Negative responses included explicit rejection notifications, as well as, more often, simply no reply.

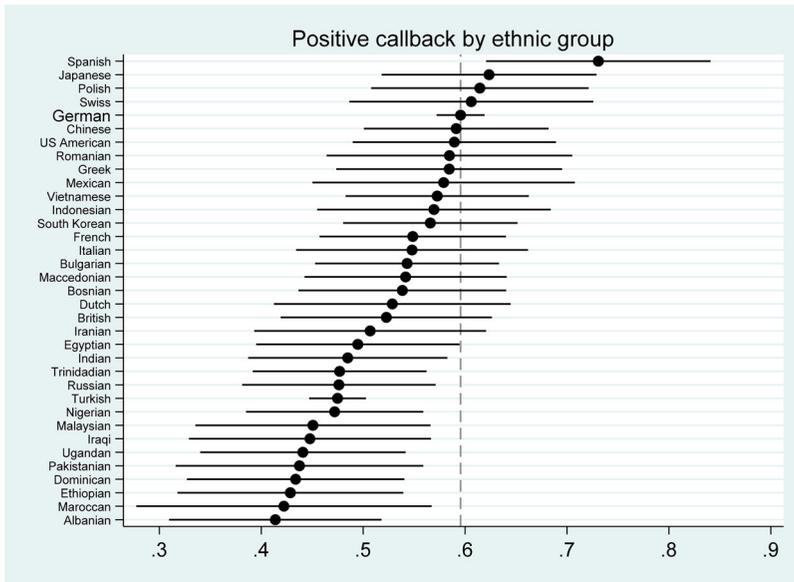
## **Results**

### ***Descriptive results***

Overall positive callback was fifty-four per cent but varied considerably across occupations (from twenty-one per cent for industrial office clerks to seventy per cent for dental assistants) and between males and females (see Online Supplement Table S1), as well as across ethnic groups. In addition, better grades have the expected positive effect, but discrimination rates do not vary significantly across grade levels. Inclusion of a reference letter of the current employer had no noticeable effect on callback.

Figure 1 shows positive callback rates – across all treatment conditions – for the 35 ethnic groups. Applicants of German origin received a positive callback in sixty per cent of the cases. Some ethnic groups had higher callback rates than German ethnics. This includes people of Spanish, Japanese, Polish and Swiss origin. Because except for German ethnics and Turkish ethnics the numbers of cases for specific ethnic groups are quite low (around  $n = 100$ ), only the Spanish result (seventy-three per cent positive callback) differs significantly from the German callback. Callback rates for Chinese, US Americans, Romanians, Greeks, Mexican, Vietnamese, Indonesians and South Koreans are above fifty-five per cent and thus less than five percentage points below those of German ethnics. All the ethnic groups from Indians (callback rate forty-eight per cent) downward differ significantly (at  $p < .05$  in linear or logistic regressions of callback on ethnicity) from German ethnics.

In this category, we also find the Turkish ethnics (callback forty-seven per cent). The absolute callback difference of thirteen percentage points (forty-seven per cent vs. sixty per cent) that we find between Turkish and German ethnics is larger than in the three previous correspondence studies of ethnic discrimination in Germany. However, in relative terms – an almost one third higher callback rate for German ethnics – our result is comparable to the results reported



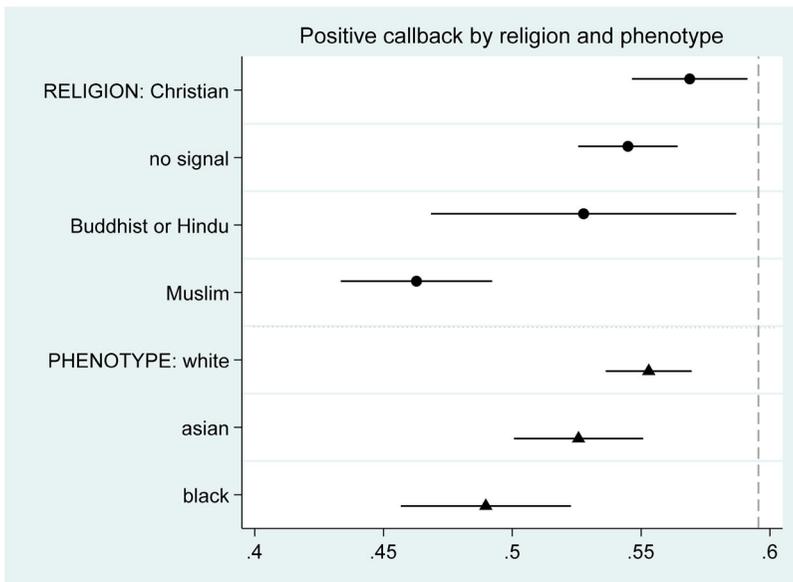
**Figure 1.** Positive callback by ethnic group.

Note: This figure shows the point estimates and confidence intervals from a linear regression of positive response on ethnicity. The dashed vertical line highlights the response rate for German job applicants without foreign roots.

by Weichselbaumer (2016, fourteen per cent vs. nineteen per cent) for Turkish and German females and by Kaas and Manger (2012, thirty-five per cent vs. forty per cent) or Schneider, Yemane, and Weinmann (2014: fifteen per cent vs. twenty per cent) for Turkish versus German males. We find that several ethnic groups have even lower callback rates than Turkish ethnics: Nigerian, Malaysian, Iraqi, Ugandan, Pakistani, Dominican, Ethiopian, Moroccan and finally Albanian ethnics, whose positive callback rate is only forty-one per cent.

Callback also varied by religion and race (see Figure 2). Positive callback for Christian applicants amounted to fifty-seven per cent and was only slightly lower for applicants without religious signal or with a Hindu or Buddhist signal (respectively fifty-four per cent and fifty-three per cent; both not significantly different from the Christian callback rate). Muslims, however, received considerably lower positive callback, at forty-six per cent (significantly different from Christians and from no religious signal at  $p < .001$ , and from Hindus or Buddhists at  $p < .10$ ). Applicants with white phenotypes had a positive callback rate of fifty-five per cent, followed by Asians with fifty-three per cent and blacks with forty-nine per cent. The difference between whites and Asians is not statistically significant but blacks have significantly lower positive callback than whites ( $p < .01$ ).

What explains these considerable differences? Are they a result of statistical discrimination and explained by productivity-related group means? Such an



**Figure 2.** Positive callback by religion and phenotype.

Note: This figure shows the point estimates and confidence intervals from two separate linear regressions of positive response on religion (see the black circles) and phenotype (see the black triangles), respectively. The dashed vertical line highlights the average response rate for German job applicants without foreign roots.

explanation has a certain degree of face validity. Ethnic groups with high positive callback rates also tend to have high mean education levels (e.g. Japanese, South Koreans, Chinese, US Americans, Swiss; see Online Supplement Table S3), and conversely many ethnic groups with low callback rates are found at the bottom of the educational ladder (e.g. Turkish, Iraqi and Moroccan ethnics). Many of these groups that combine low education levels with low positive callback are predominantly Muslim. However, alternative taste-based explanations also have face validity. The West European and East Asian ethnic groups with very high callback rates originate in countries that are very close to Germany in terms of emancipative values (again, see Table S3). And conversely, ethnic groups that have low callback such as Pakistani, Nigerian, Turkish, Ugandan and Moroccan ethnics have values very different from those prevalent in Germany. Black phenotypes and Muslim religiosity are disproportionately concentrated in these groups that combine high value distance with low positive callback.

### **Multivariate results**

Thanks to the design of our study, we can move beyond speculation and investigate which of the two theories of discrimination has the greatest

explanatory power. To this end, we ran a series of regression models. Because our observations are nested in altogether 132 different combinations of ethnic groups, phenotypes and religious signals, we fitted linear mixed-effects models with random intercepts by these ethnicity-phenotype-religion clusters (see Gelman and Hill 2006; McCulloch and Neuhaus 2011). Following Hellevik (2009), we ran linear instead of logistic regressions, which have the advantage that coefficients are easily interpretable and comparable across models. Results of alternative model specifications are discussed below in the robustness checks section and shown in Online Supplement Tables S5–S6.

The first column of Table 1 shows the baseline model, which includes the ethnic origin, racial phenotype and religion treatments as well as all fully orthogonal treatments and controls (grades, reference letter, gender, occupation and month of application). Turkish ethnics have a nine percentage points lower likelihood of receiving a positive response than German ethnics. Other non-German ethnics are – on average – less disadvantaged than Turkish ethnics, with a response rate that is five percentage points below the rate for German ethnics. Applicants with black phenotypes have a positive callback rate that is seven percentage points below the rate for whites. Finally, for applicants volunteering for a Muslim organization the likelihood of a positive response is seven percentage points lower than for Christian volunteers. All these results are statistically significant. Except for a significantly weaker penalty of signalling Muslim faith among Turkish immigrant compared to other immigrants, there are no significant interactions between the three types of penalties – ethnic, racial and religious (see Online Supplement Table S9). The added effects of disadvantaged traits cumulate in the group of black Muslims of immigrant origin, which the model predicts to have nineteen percentage points lower callback rates (see the coefficients of  $-.05$  for other non-German ethnics,  $-.07$  for blacks and  $-.07$  for Muslims).

In column 2, we test the statistical discrimination hypothesis and investigate to what extent education levels as an indicator of average group productivity can explain the observed penalties. The aim of these analyses is not to increase the overall explained variance. What we are looking for in these models is whether indicators of group productivity and values absorb or diminish the effects of ethnic, racial and religious group membership signals. In line with hypothesis 1, we find a significant effect of average education on positive callback. The coefficient of  $.03$  implies that a one standard deviation increase on the education variable (.21 points on the education scale) results in 3.5 per cent higher callback. While the education variable reduces the size of the Muslim treatment effect from  $-.07$  in model 1 to  $-.04$  in model 2 and the effect of Turkish ethnicity from  $-.09$  to  $-.04$ , it does not reduce the size of the penalty for other non-German ethnics and for black applicants. The reason is that Turkish ethnics and groups originating

**Table 1.** Linear regressions of positive callback on individual-level treatments and group-level covariates.

	<i>Main analyses (ethnicity–religion–phenotype clusters)</i>				<i>Simplified analyses (ethnicity clusters)</i>			
	<i>Model 1: Baseline</i>	<i>Model 2: Education</i>	<i>Model 3: Value distance</i>	<i>Model 4: Education and value distance</i>	<i>Model 5: No group-level covariates</i>	<i>Model 6: Education</i>	<i>Model 7: Value distance</i>	<i>Model 8: Education and value distance</i>
<b>Ethnicity (ref.: German)</b>	-.09 (.02)***	-.04 (.03)	-.00 (.03)	.01 (.03)	-.09 (.02)***	-.05 (.03)*	-.01 (.03)	.01 (.03)
Other non-German	-.05 (.02)**	-.05 (.02)***	.01 (.02)	-.01 (.03)	-.05 (.02)**	-.05 (.02)***	.01 (.02)	-.01 (.03)
<b>Phenotype (ref.: white)</b>	-.07 (.02)***	-.08 (.02)***	-.02 (.02)	-.04 (.03)	-.07 (.02)***	-.07 (.02)***	-.06 (.02)**	-.07 (.02)**
Black	-.01 (.02)	-.03 (.02) <sup>+</sup>	.01 (.02)	-.00 (.02)	-.01 (.02)	-.02 (.02)	-.00 (.02)	-.02 (.02)
Asian	-.01 (.02)	-.01 (.02)	-.00 (.02)	-.00 (.02)	-.01 (.02)	-.01 (.02)	-.01 (.02)	-.01 (.02)
<b>Religion (ref.: Christian)</b>								
No Muslim	-.07 (.02)**	-.04 (.02) <sup>+</sup>	-.03 (.02)	-.03 (.02)	-.07 (.02)**	-.06 (.02)**	-.05 (.02)*	-.04 (.02) <sup>+</sup>
Hindu or Buddhist	-.02 (.03)	-.05 (.03)	.00 (.03)	-.01 (.04)	-.02 (.03)	-.03 (.03)	-.02 (.03)	-.03 (.03)
<b>Education level</b>		.03 (.01)**		.01 (.01)		.02 (.01) <sup>+</sup>		.01 (.01)
<b>Value distance</b>			-.04 (.01)***	-.03 (.01)*			-.03 (.01)**	-.03 (.01)*
<b>CONTROLS VARIABLES:</b>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N (groups)	132	132	132	132	35	35	35	35
N (applications)	5,819	5,819	5,819	5,819	5,819	5,819	5,819	5,819
R <sup>2</sup> (within)	.08	.08	.08	.08	.08	.08	.08	.08
R <sup>2</sup> (between)	.17	.21	.22	.22	.47	.49	.55	.55
R <sup>2</sup> (overall)	.09	.09	.09	.09	.09	.09	.09	.09

Notes: Table 1 shows the results of random-intercept linear regression models with standard errors clustered at the level of 132 ethnicity–religion–race clusters (models 1–4) and 35 ethnic groups (models 5–8). Group-level covariates are standardized. All models include the following control variables and orthogonal treatments: grades, reference letter, gender, occupation and month of application.

Standard errors in parentheses; *p*-levels: <sup>+</sup>*p* < .10, \**p* < .05, \*\**p* < .01, \*\*\**p* < .001 (one-tailed).

in Muslim countries more generally indeed have relatively low education levels. Immigrants from non-Muslim European, Asian and African countries, however, have quite high average education levels. As a result, education levels provide a potential explanation for discrimination against Turks and other Muslims, but not against other immigrant groups.

Next, we test the idea derived from taste-based theories of discrimination that preferences related to cultural values and norms shape patterns of discrimination. The third column of [Table 1](#) shows that value distance has a strong impact on discrimination rates. The coefficient of  $-0.04$  implies that a one standard deviation increase in value distance from the white-Christian-German baseline results in 4.9 per cent lower callback. In this model, all ethnic, racial and religious penalties are strongly reduced and become statistically insignificant. We, therefore, find strong support for hypothesis 2 on taste discrimination based on cultural value differences.

In a final step, we test hypotheses 1 and 2 simultaneously. The results of model 4 show that the effect of value distance is only slightly reduced and remains significant, whereas the education effect is strongly reduced and becomes insignificant. Mediation analyses confirm that value-based taste discrimination explains the pattern of ethnic discrimination much better than education-based statistical discrimination does (see Online Supplement Table S11).

In models 5–8 of [Table 1](#) we report results when we use simplified measures of education levels and value distance that are only calculated across the thirty-five ethnic groups. Because these do not vary within ethnic groups across phenotypical and religious signals, they cannot explain the effects that indicate discrimination based on these signals. Regarding ethnicity, however, they confirm the results of the main analysis: education levels do not explain much of the discrimination against Turkish and other non-German ethnics, whereas value distance explains ethnic discrimination very well and absorbs the effects of Turkish and other non-German ethnicity almost entirely.

### **Robustness checks**

To demonstrate robustness, we replicated all multivariate analyses with different model specifications (see the Online Supplement):

- Using random-intercept logistic regression to account for the binary dependent variable, we found no meaningful differences anywhere (see [Table S5](#)).
- We weighted all thirty-four non-German ethnic groups equally in order to remove potential biases resulting from the larger number of applicants of Turkish origin, as well as the smaller variations in sample size across the

other groups. These results do not deviate in any substantively important way either (see Table S6).

- Next, we further scrutinized the result that group-level value distance trumps education levels by splitting the sample in half in three different ways: male versus female applicants; jobs with low and high customer contact; and jobs with a lower and a higher required level of schooling. In each of the six regressions, the magnitude of the group-level coefficients remains virtually unchanged and value distance largely absorbs the ethnicity effect (see Table S7).
- We also considered other indicators of productivity, namely the average unemployment rates and occupational status of groups. The alternative productivity indicators explain the pattern of discrimination more or less equally well as value distance. This implies that we cannot reject statistical discrimination explanations with certainty. Nevertheless, the evidence we present for taste discrimination is stronger because it is based on an indicator – value distance – that is independent of labour-market outcomes, and because our statistical discrimination indicator that is less strongly plagued by problems of reverse causality – education levels – performs poorly in comparison (see Table S8).
- In a final step, we tested an alternative model that instead draws on productivity information on the individual level. In Table S10, we added interaction terms between grades and ethnicity, phenotype and religion, respectively, to our main model. None of the interaction terms had a significant coefficient. This suggests that while employers generally value productivity information, majority and minority members benefit equally from better grades.

## Discussion and conclusions

By employing a novel research design that allowed comparing the impact on labour-market discrimination of various ethnic, racial and religious signals simultaneously, we have shed light both on the relative importance of these signals, and on the key theoretical question whether discrimination is driven by statistical or taste considerations. Like many previous studies, ours reveals significant discrimination against minority applicants. However, this discrimination does not affect all minority groups equally and some of them not at all. While we found strong discrimination against Muslims, there was no significant bias against Buddhists or Hindus. While blacks were significantly discriminated against, this was much less the case for Asian applicants.

We identified taste-based discrimination against minority groups with values that differ strongly from the German value pattern as the most

important factor explaining these differences. Once value distance is considered, we no longer find statistically significant penalties against applicants with non-German ethnic origins, non-white phenotypes and non-Christian religiosity. This result indicates that there is no significant discrimination against minorities that are culturally proximate to Germans. This accounts for the high positive callback rates of groups of West European, South European and US American origin, but also of those of East Asian origin, who all have value patterns close to Germany's. The value patterns of Sub-Saharan African and Muslim groups are the most distant from Germans and they are also found at the bottom of the callback ranking.

While positive callback was also correlated with mean group levels of education, our indicator of statistical discrimination, this variable could not account for most of the group differences in callback and its effect was much reduced and became statistically insignificant when value distance was considered simultaneously. Once we control for value distance, we no longer observe significant ethnic, phenotypical and religious penalties. Value distance thus offers a satisfactory explanation for the patterns of discrimination found in the data.

Our results suggest that employers' decisions are sensitive to group stereotypes regarding attitudes towards freedom, autonomy and gender equality. This is in line with Adida, Laitin, and Valfort (2016) who report that French employers referred to conflicts related to such values – e.g. over relations with co-workers of the opposite sex – when asked about reasons behind discrimination against Muslims. Whether employers' reference to social values is purely a matter of taste or whether it is also related to actual productivity concerns relating to cultural conflicts on the job cannot be answered with our research design. Even if value concerns can have a basis in empirical fact on the level of the average social values of groups, it goes without saying that discrimination against individuals because of assumptions about the average traits of the groups they belong to is unethical and, in most countries, illegal. However, the more subtle forms of taste discrimination that our results reveal may be more difficult to identify and combat than crude taste discrimination based on ethnicity, race, or religion per se.

Culture and values would be a good starting point for future research to investigate employer rationales behind discrimination. We hope to inspire others to use the kind of multiple-group design that we have employed to answer the question to what extent these results generalize beyond Germany. Application materials in Germany are extremely detailed and include copies of all relevant certificates and diplomas. Given this rich individual information, there may be less need for German employers to rely on group-level indicators of productivity than in other countries (similarly Zschirnt and Ruedin 2016). The relative importance of racial and religious signals may also differ in other contexts, and ethnic hierarchies in other

immigration countries may deviate from those in Germany. Race could, for instance, be a more salient signal in countries with a history of slavery and colonialism, whereas the weight of religious signals might be related to the salience of debates on Islam in the European context.

## Notes

1. For a few small ethnic groups, no separate entries are available in the Mikrozensus and their values had to be approximated by those of larger geographical categories. For instance, the Egyptian values are taken from the grouping “Egypt, Algeria, Libya, and Tunisia”. Of course, these are rough approximations, but this data are the most detailed data that are available in Germany.
2. In the main analyses we focus on average education levels. However, we consider groups’ average unemployment rate and occupational status (ISEI) as alternative indicators of productivity. We have chosen not to use these alternative indicators for our main analysis because unemployment rates and occupational status may themselves be a result of labor market discrimination and therefore the direction of causality is unclear. Online Supplement Table S8 shows the results of analyses with these alternative indicators.
3. Country-averages provide imperfect estimates for cultural values among different immigrant groups, but they are the best proxies that we are aware of. Country-averages of cultural values may differ from employer stereotypes about immigrant groups, for example, in consequence of self-selection into migration (Chiswick 1999) or value changes after migration (Lönngqvist, Jasinskaja-Lahti, and Verkasalo 2011).
4. Greeks receive the same values as their neighbors, the Macedonians, because the value surveys provide no information about Greeks.
5. Information about the prevalence of different religious groups across countries is provided by the Pew research center (<http://www.pewforum.org/>).
6. Some countries, especially in Central and South America, are difficult to classify because most of the population is of mixed race. We have left these countries out of the calculation of the phenotype average scores. The boundary between predominantly white and predominantly Asian phenotypes (which we draw between Iran and Pakistan) is to some extent arbitrary, but results are robust to alternative specifications. The countries of Sub-Saharan Africa are designated as those with a predominantly black phenotype.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Adida, Claire L., David D. Laitin, and Marie-Anne Valfort. 2016. *Why Muslim Integration Fails in Christian-Heritage Societies*. Cambridge: Harvard University Press.
- Agerström, Jens, Fredrik Björklund, Rickard Carlsson, and Dan-Olof Rooth. 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.

- Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labour Markets." *Industrial and Labour Relations Review* 30 (2): 175–187.
- Arrow, Kenneth J. 1973. "The Theory of Discrimination." *Discrimination in Labour Markets* 3 (10): 3–33.
- Baert, Stijn, Bart Cockx, Niels Gheyle, and Cora Vandamme. 2015. "Is There Less Discrimination in Occupations Where Recruitment Is Difficult?" *ILR Review* 68 (3): 467–500.
- Baert, Stijn, and Ann-Sophie De Pauw. 2014. "Is Ethnic Discrimination Due to Distaste or Statistics?" *Economics Letters* 125 (2): 270–273.
- Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." *The American Economic Review* 106 (6): 1437–1475.
- Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Bertrand, Marianne, and Esther Duflo. 2017. "Field Experiments on Discrimination." *Handbook of Economic Field Experiments* 1: 309–393.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labour Market Discrimination." *American Economic Review*: 94: 991–1013.
- Blommaert, Lieselotte, Marcel Coenders, and Frank Van Tubergen. 2014. "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.
- Booth, Alison L., Andrew Leigh, and Elena Varganova. 2012. "Does Ethnic Discrimination Vary Across Minority Groups? Evidence From a Field Experiment: Discrimination Across Minority Groups." *Oxford Bulletin of Economics and Statistics* 74 (4): 547–573.
- Byrne, Donn. 1997. "An Overview (and Underview) of Research and Theory Within the Attraction Paradigm." *Journal of Social and Personal Relationships* 14 (3): 417–431.
- Chiswick, Barry. 1999. "Are Immigrants Favorably Self-Selected?" *American Economic Review* 89 (2): 181–185.
- Derous, Eva, Ann Marie Ryan, and Hannah-Hanh D. Nguyen. 2012. "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.
- Drydakis, Nick. 2010. "Religious Affiliation and Employment Bias in the Labour Market." *Journal for the Scientific Study of Religion* 49 (3): 477–493.
- Edo, Anthony, Nicolas Jacquemet, and Constantine Yannelis. 2013. "Language Skills and Homophilous Hiring Discrimination: Evidence From Gender- and Racially-Differentiated Applications." CES Working Papers 2013-58. Paris: CES.
- England, Paula, and Peter Lewin. 1989. "Economic and Sociological Views of Discrimination in Labour Markets: Persistence or Demize?" *Sociological Spectrum* 9 (3): 239–257.
- EVS. 2015. "European Values Study Longitudinal Data File 1981–2008 (EVS 1981–2008)." GESIS Data Archive, Cologne. ZA4804 Data file Version 3.0.0. doi:10.4232/1.12253.
- Gaddis, S. Michael. 2017. "How Black Are Lakisha and Jamal? Racial Perceptions From Names Used in Correspondence Audit Studies." *Sociological Science* 4: 469–489.
- Gaddis, S. Michael. 2018. "An Introduction to Audit Studies in the Social Sciences." In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, edited by S. Michael Gaddis, 3–44. Cham: Springer. [E-book] <https://ebookcentral.proquest.com/>.

- Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Guryan, Jonathan, and Kerwin Kofi Charles. 2013. "Taste-Based or Statistical Discrimination: The Economics of Discrimination Returns to Its Roots." *The Economic Journal* 123 (572): F417–F432.
- Hellevik, Ottar. 2009. "Linear Versus Logistic Regression When the Dependent Variable Is a Dichotomy." *Quality & Quantity* 43 (1): 59–74.
- Jacquemet, Nicolas, and Constantine Yannelis. 2012. "Indiscriminate Discrimination: A Correspondence Test for Ethnic Homophily in the Chicago Labour Market." *Labour Economics* 19 (6): 824–832.
- Kaas, Leo, and Christian Manger. 2012. "Ethnic Discrimination in Germany's Labour Market: A Field Experiment." *German Economic Review* 13 (1): 1–20.
- Lönnqvist, Jan-Erik, Inga Jasinskaja-Lahti, and Markku Verkasalo. 2011. "Personal Values Before and After Migration: A Longitudinal Case Study on Value Change in Ingrian–Finnish Migrants." *Social Psychological and Personality Science* 2 (6): 584–591.
- McCulloch, Charles E., and John M. Neuhaus. 2011. "Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter." *Statistical Science* 26 (3): 388–402.
- McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27 (1): 415–444.
- Montoya, R. Matthew, Robert S. Horton, and Jeffrey Kirchner. 2008. "Is Actual Similarity Necessary for Attraction? A Meta-Analysis of Actual and Perceived Similarity." *Journal of Social and Personal Relationships* 25 (6): 889–922.
- Neumark, David. 2016. "Experimental Research on Labour Market Discrimination." Working Paper 22022. National Bureau of Economic Research.
- Oreopoulos, Philip. 2011. "Why Do Skilled Immigrants Struggle in the Labour Market? A Field Experiment with Thirteen Thousand Resumes." *American Economic Journal: Economic Policy* 3 (4): 148–171.
- Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108 (5): 937–975.
- Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *The ANNALS of the American Academy of Political and Social Science* 609: 104–133.
- Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *The American Economic Review* 62 (4): 659–661.
- Pienné, Guillaume. 2013. "Hiring Discrimination Based on National Origin and Religious Closeness: Results From a Field Experiment in the Paris Area." *IZA Journal of Labour Economics* 2 (1): 4. doi:10.1186/2193-8997-2-4.
- Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. 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. doi:10.1073/pnas.1706255114.
- Rich, Judith. 2014. "What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted Since 2000." IZA Discussion Papers 8584. Available at SSRN. <https://ssrn.com/abstract=2517887>
- Saeed, Abubakr, Sidra Maqsood, and Amir Rafique. 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. doi:10.1080/13547860.2019.1612974.

- Schneider, Jan, Ruta Yemane, and Martin Weinmann. 2014. *Diskriminierung am Ausbildungsmarkt. Ausmaß, Ursachen und Handlungsperspektiven*. Berlin: Sachverständigenrat Deutscher Stiftungen für Integration und Migration.
- Veit, Susanne, and Ruta Yemane. 2018. "The ADIS Study: A Large-scale Correspondence Test on Labour Market Discrimination in Germany: Technical Report". WZB Discussion Paper SP VI 2018-103.
- Wallace, Michael, Bradley RE Wright, and Allen Hyde. 2014. "Religious Affiliation and Hiring Discrimination in the American South: A Field Experiment." *Social Currents* 1 (2): 189–207.
- Weichselbaumer, Doris. 2016. "Discrimination Against Female Migrants Wearing Headscarves." IZA Discussion Papers, No. 10217.
- Welzel, Christian. 2013. *Freedom Rising*. New York: Cambridge University Press.
- Wright, Bradley R. E., Michael Wallace, John Bailey, and Allen Hyde. 2013. "Religious Affiliation and Hiring Discrimination in New England: A Field Experiment." *Research in Social Stratification and Mobility* 34 (1): 111–126.
- WVS. 2015. "World Value Survey 1981-2014 Official Aggregate v.20150418, 2015. World Values Survey Association ([www.worldvaluessurvey.org](http://www.worldvaluessurvey.org))." Aggregate File Producer: JDSYSTEMS, Madrid.
- 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 (7): 1115–1134.