

DOUBLE JEOPARDY UPON RESUMÉ SCREENING: WHEN ACHMED IS LESS EMPLOYABLE THAN AÏSHA

EVA DEROUS
Ghent University

ANN MARIE RYAN
Michigan State University

ALEC W. SERLIE
GITP

Applicants belong to multiple categories (e.g., male, ethnic minority) and a complex set of factors affects category activation and inhibition when making hiring decisions. Two field experiments with recruiters who regularly engage in resumé screening showed that the role of multiple categories (applicants' ethnicity and sex) in discrimination depended on job type and prejudice. Specifically, in both low- and high-demand (i.e., complex) jobs, Arab women were rated more favorably than Arab men, particularly when considering levels of client contact. Across both studies, recruiters high in explicit ethnic prejudice were discriminatory only when applicants' job qualifications fit the job position less, lending support for the attributional-ambiguity effect. Implicit attitudes did not play a strong role. Our study findings point to the complex nature of multiple categorization effects in the hiring process. Implications are considered as to how to avert hiring discrimination during resumé screening.

Resumés are one of the most important sources of information that personnel decision-makers consider when they initially screen applicants for jobs, but resumé screening may be highly susceptible to cognitive bias. Models of impression formation, such as the continuum model (Fiske, Lin, & Neuberg, 1999) and the dual process model (Brewer & Harasty Feinstein, 1999), suggest that category-based information processing occurs automatically and will be particularly strong when limited information about an individual is available. In resumé screening, when job applicants are judged on the basis of a one- or two-page resumé only,

The authors recognize Beatrijs Moerkerke, Jan Lammertyn, Jan De Houwer, and Maarten De Schryver for their helpful comments on prior versions of this article. Partial funding for this study came from the SIOP Foundation via the Small Grant Program.

Correspondence and requests for reprints should be addressed to Eva Deros, Department of Personnel Management, Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium; eva.derous@ugent.be.

information about applicants is limited and judgment may be based on raters' category-based stereotypes.

Although information processing based on social category membership (e.g., race, sex) is a well-documented phenomenon in resumé screening (Bertrand & Mullainathan, 2004; Cole, Rubin, Field, & Giles, 2007; De Beijl, 2000; Derous, Nguyen, & Ryan, 2009), little is known about how multiple social categories are considered in conjunction when making hiring decisions (Kulik, Roberson, & Perry, 2007). In particular, debate exists as to whether applicant sex moderates ethnicity effects (Cleveland, Vescio, & Barnes-Farrell, 2005), with some researchers suggesting that minority women are the target of the greatest discrimination (Berdahl & Moore, 2006), others suggesting minority men fare worse in hiring evaluations (Sidanius & Pratto, 1999), and still others adopting the perspective that who is most affected depends upon the job (De Beijl, 2000; Holzer & Inhlantfeldt, 1998). In summarizing this literature, Kulik et al. (2007) provided a series of propositions related to category activation and inhibition in hiring processes that remain untested. The first contribution of this study is to examine these competing views regarding multiple category membership in the context of resumé screening, based on the propositions of Kulik et al. (2007).

Second, Kulik et al. (2007) specifically note that social categorization effects are likely impacted by contextual and individual influences, but to date these have not been considered in evaluating multiple category membership effects in hiring contexts. We contribute to understanding of multiple categorization effects by examining potential moderators of such effects: job context, rater individual differences, and category salience. In doing so, we are able to provide a more complete, albeit more complex, examination of multiple social categorization in hiring contexts.

Many studies on resumé screening are conducted with student raters instead of actual recruiters (e.g., Blommaert, van Tubergen, & Coenders, 2012; Ziegert & Hanges, 2005). Student samples can provide evidence but might be less than ideal if participants are uninterested or inexperienced (Landy, 2008). As a third contribution, we conducted two field experiments among *actual* recruiters to enhance the ecological validity of study findings.

We first discuss competing theoretical perspectives on multiple category membership effects on evaluations of individuals. We then describe a study to evaluate two of the potential influences on those effects: job characteristics and rater prejudice. In a second study, we examine ethnicity salience along with an alternative manipulation of job characteristics and additional measures of prejudiced attitudes. While our aim is to produce understanding of multiple categorization effects that has broad generalizability across hiring contexts, we recognize that examinations of

discrimination are always bounded by societal context (Booth, Leigh, & Varganova, 2012) and the particular social groups targeted. The focus of these studies is on Arabs in the Netherlands, where they are the largest ethnic minority group with unemployment rates twice as high as those of host nationals and other ethnic minorities (Andriessen, Nievers, & Dagevos, 2012). Thus, examination of discrimination in resumé screening for this group is analogous to a number of other societal contexts where ethnic minorities fare worse in the job market than ethnic majorities (Organization for Economic Co-operation and Development [OECD], 2008).

Multiple Categorization

Individuals belong to multiple social groups (e.g., ethnic, sex, religion, age, sexual orientation) and have multiple social identities (Ashkanasy, Härtel, & Daus, 2002; Nelson & Probst, 2004). Two prominent hypotheses as to how multiple social categories might affect evaluations are the *double jeopardy hypothesis* and the *subordinate male target hypothesis* (Sidanius & Veniegas, 2000). Both hypotheses predict that considering the two social categories together (ethnicity and sex) provides greater insight into the nature of discrimination than studying them in isolation (Berdahl & Moore, 2006; Sidanius & Veniegas, 2000).

According to the double jeopardy hypothesis, ethnic minority women would suffer more hiring discrimination than ethnic majority women and majority/minority men. That is, the double jeopardy hypothesis presents minority ethnicity and women as two outgroup categories where an individual who is a member of both will experience more negative evaluations, in either an additive or multiplicative fashion, than would those belonging to only one outgroup (majority women; minority men) or no outgroups (majority men; see for an example in the context of sexual harassment: Berdahl & Moore, 2006). The subordinate male target hypothesis also indicates attention to both categories but suggests that minority men would be evaluated the most negatively, as they are the most threatening group. These symbolic threats can be economic, political, or monetary, such as threatening employment security or advancement opportunities (as per realistic group conflict theory; Esses, Jackson, & Armstrong, 1998). Although the double jeopardy hypothesis is more often discussed in the literature, there is some support for the subordinate male target hypothesis upon resumé screening (e.g., Bendick, Jackson, & Reinoso, 1994; Bendick, Jackson, Reinoso, & Hodges, 1991; Ghumman & Jackson, 2010; Hosoda, Stone, & Stone-Romero, 2003). For instance, a review of employment audit studies (Sidanius & Pratto, 1999) showed that the level of employment discrimination against minority men (29.5%) was substantially larger than the employment discrimination against minority women

(22.5%) after controlling for human capital factors like required skill level for the job.

The stereotype content model (Cuddy, Fiske, & Glick, 2007) suggests that groups may be discriminated against in different ways, depending upon the content of the stereotype of the group. In empirical research on this model, Arabs are evaluated as low in both warmth and competence, a stereotype content that has been connected to feelings of contempt and greater active and passive harm behaviors from other groups. However, gender differences in stereotypes also exist. In the Netherlands, labor market outcomes of Arab men are generally lower than those of Arab women, and employers' attitudes toward Arab men are more negative (e.g., being perceived as more hostile) compared to Arab women (e.g., being perceived as more submissive; Andriessen, Nievers, Faulk, & Dagevos, 2010; Blom, Oudhof, Bijl, & Bakker, 2005; Dagevos, 2011; Klaver, Mevissen, Odé, Materman, & Weening, 2005; OECD, 2013), pointing in the direction of the subordinate male target hypothesis in the specific context of our study. Kulik et al. (2007) further argue that in hiring contexts with limited information, managers are likely to use less effortful processing (Macrae & Bodenhausen, 2000) and will not consider multiple categories simultaneously but let one dominate. Given these findings, the subordinate male hypothesis is more likely than the double jeopardy hypothesis despite the latter's greater popularity. Therefore, we propose:

Hypothesis 1: Consistent with the subordinate male target hypothesis, job suitability ratings will be lower for resumés of Arab men compared to equally qualified Arab female, Dutch female, and Dutch male applicants.

Job Contingencies

Kulik et al. (2007) also suggest that neither of the perspectives of cumulative effects (double jeopardy or subordinate target male hypotheses) might be correct. A matching model might predominate where category memberships and accompanying stereotypes are considered in relation to specific job requirements (e.g., the job requires assertiveness and a group is stereotyped as submissive or aggressive). Kulik et al. note that decision makers may face contexts where stereotypes associated with one category (e.g., Asian as nonassertive) might conflict with stereotypes of another category (e.g., men as assertive), in which case, category inhibition or activation will determine how the decision maker resolves this "multiple category problem" (i.e., whether an Asian man is evaluated as a good fit). In order to consider this possibility, we examined how the nature of

the job might affect views of multiple category membership in resumé screening.

One job characteristic that might play an important role is the amount of external client interaction required in the position. For instance, De Beijl (2000) found that migrant applicants' rejection rates were significantly higher in jobs that required visual contact with clients. Holzer and Ihlanfeldt (1998) showed the lower the fraction of minority customers, the higher the probability of hiring discrimination against applicants from that minority group. Majority group clients may perceive outgroup staff as less similar to them, and that may negatively affect clients' perception of an organization (Feagin & Eckberg, 1980). By hiring few visible ethnics in high client-contact positions, employers may feel they are safeguarding the organization against clients' potential negative reactions.

However, De Beijl (2000) and others (e.g., Deros et al., 2009) only investigated effects of client contact among male applicants; it remains unclear whether these findings also hold for female minority applicants. The nature of the job might influence whether those in certain combinations of social categories are seen as a match. As Kulik et al. (2007) noted, "job applicants categorized as members of negatively evaluated categories can still be positively evaluated if they match the perceived requirements of the job" (p. 536). For instance, in the Netherlands, employers may perceive Arab minority women as more docile and helpful than their male counterparts (Dagevos, 2011), and therefore, they may be seen as fitting service-related positions. In the case of our study context, the matching hypothesis would suggest that Arab women would be viewed as a better match for positions requiring high client contact because of stereotypes of Arab women as helpful. This goes beyond Hypothesis 1 in that the subordinate male target effect is expected to hold for high- and low-contact jobs (i.e., with external client contact). However, minority men will be evaluated lower in high-contact jobs because the stereotype-position match is more salient for Arab women.

Hypothesis 2: Contact with clients will moderate the effect of applicants' minority status on job suitability ratings such that the difference between Arab female and male ratings will be greater for high client contact positions than for low client contact positions.

Recruiters' Prejudice

Kulik et al. (2007) also noted that attitudes toward members of groups (i.e., prejudice) will influence category activation and inhibition. Prejudice has been well established as a predictor of engagement in stigmatizing

behavior (Dovidio & Hebl, 2005; Macrae & Bodenhausen, 2000) and can operate in subtle and implicit ways (Brief, Dietz, Cohen, Pugh, & Vaslow, 2000). There is some evidence that ethnic minorities receive lower hiring outcomes when recruiters score high on prejudiced attitudes (Blommaert et al., 2012). Yet, some mixed findings have been reported too: Both Derous et al. (2009) and Son Hing, Chung-Yan, Hamilton, and Zanna (2008) found no effect of ethnic prejudice on hiring outcomes. Furthermore and to the best of our knowledge, no research has ever considered recruiters' sexism, measured in less overt terms, upon resumé screening.

Kulik et al. (2007) proposed that in the case of multiple categories, the category with the strongest attitudes will be activated and dominate the decision process. In other words, those high in ethnic prejudice will pay more attention to ethnicity and those high in sexism will pay more attention to gender as these categories are more highly accessible for prejudiced individuals. To further investigate potential moderating effects of ethnic prejudice (explicitly measured), we hypothesize that:

Hypothesis 3: Raters high in explicit ethnic prejudice toward Arabs will give lower job suitability ratings to Arabs than those low in explicit prejudice (Hypothesis 3a), and those high in explicit sexism will give lower job suitability ratings to women than low explicit sexism raters (Hypothesis 3b).

The most common way to measure prejudiced attitudes, through explicit self-report measures, may be vulnerable to self-presentation biases (Talaska, Fiske, & Chaiken, 2008; Uhlmann, Dasgupta, Elgueta, Greenwald, & Swanson, 2002). Researchers have directed their attention to more *implicit* ways of measuring prejudiced attitudes, such as the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The principle behind the IAT is that tasks that are consistent with one's social-cognitive associations will be performed in a more fluent and efficient way than those that are inconsistent with one's social-cognitive associations (Rudman, 2008). A recent meta-analysis (Greenwald, Poehlman, Uhlmann, & Banaji, 2009) showed stronger predictive utility for IATs than for explicit prejudice measures, particularly so for socially sensitive topics (like ethnic discrimination). IAT measures have been used and validated in different psychological domains, such as social behaviors (e.g., Fazio & Olson, 2003; Greenwald et al., 2009; Kihlstrom, 2004, for reviews), consumer behaviors (e.g., Brunel, Tietje, & Greenwald, 2004), and clinical disorders (e.g., de Jong, van den Hout, Rietbroek, & Huijding, 2003).

Despite some criticism (e.g., Blanton & Jaccard, 2006a, 2006b; Landy, 2008), IATs have also been postulated as a very promising tool

in organizational research (Haines & Sumner, 2006) but not investigated extensively. Recently, effects on “hireability ratings” have been reported by a few authors in lab (Blommaert et al., 2012; Derosus et al., 2009; Rudman & Glick, 2001) and field settings (Rooth, 2010). Kulik et al. (2007) specifically noting the need to use non-self-report measures in studying multiple category effects because category activation and inhibition are not always conscious processes. In keeping with their suggestion, we also measured prejudice using an implicit measure:

Hypothesis 4: Raters high in implicit prejudice toward Arabs will give lower job suitability ratings to Arabs than those low in implicit prejudice toward Arabs (Hypothesis 4a) and those high in implicit sexism will give lower job suitability ratings to women than those low in implicit sexism (Hypothesis 4b).

We also explored two additional questions related to the role of prejudice. First, if socially desirable responding leads to less accuracy in explicit prejudice measures, we might expect stronger interaction effects for the implicit prejudice measures than the explicit ones. The possibility of differential effects for implicit and explicit measures was explored. Second, one reason why findings regarding prejudice’s effects on hireability ratings (whether measured explicitly or implicitly) are not always consistent is that the job context may influence whether an individual acts on his/her prejudice (Heilman & Haynes, 2008). Prejudiced individuals may only engage in discriminatory acts in situations with high attributional ambiguity. That is, prejudice influences behavior when there exist nonethnic-related justifications for discriminatory reactions toward ethnic minorities. Examples of justifications include the amount of external client contact (Brief et al., 2000; Ziegert & Hanges, 2005) or when there is not a perfect match between the job requirements and applicants’ qualifications (Dovidio & Gaertner, 2000; Hodson, Dovidio, & Gaertner, 2002; Son Hing et al., 2008). Thus, we also explored whether the effect of recruiters’ ethnic/gender prejudiced attitudes (either explicitly or implicitly measured) on hireability ratings were contingent upon the job context.

Study 1

Method

Participants. Respondents were 60 non-Arab/Dutch recruiters (45.1% males) identified through (a) professional networks of several HR-professionals in the Netherlands, their consultants, commissioning clients,

and professional contacts, as well as (b) the HR-network of the business school at the university. To stimulate recruiters to participate, we offered them a book on HR (value: €25.00), and in addition, we organized a lottery with vouchers (value: €20.00), which recruiters could freely spend in a shop of their choice. All participants were experienced in recruiting applicants and had between 3 and 10 years of work experience (Table 1).

Development of materials and pilot tests. Study materials (resumés/advertisements/IATs) were developed, and separate pilot studies were conducted to ensure equivalence.¹ First, names (e.g., Mohammed, Henk) and affiliations (e.g., member of Arab association, member of rowing club) were evaluated as to Arab or Dutch ethnicity ($n = 91$), and names and affiliations were chosen for use in the resumés based on strength of association (see the Appendix). Next, the cognitive demands (*low vs. high*), the amount of face contact with external clients (*low vs. high*), and the gendered nature of work (*male-dominated job vs. female-dominated job*) of 34 jobs were evaluated ($n = 111$). Based on the ratings, we selected five jobs within the service/consultancy sector that differed in level of cognitive demands and external client contact but that were equally accessible to men and women and required a similar educational background. Thereafter, we prepared advertisements for each job, which served as the manipulation of client contact. For Study 1, two clerical jobs were selected, one with little external client contact and one with frequent external client contact. For Study 2, we selected three consultant jobs with high cognitive demands that differed in amount of external client contact (*little, moderate, high*). We then ensured relevance and equivalency of work experiences ($n = 141$). For each applicant profile (resumé), we selected two work experiences, one with much external client contact and one with less external client contact. Finally, participants (same $n = 91$ from the first pilot) were asked to compare four resumés (per job type) in a pair-wise order to judge the overall equivalence of nonmanipulated resumé aspects (e.g., education, amount of work experience) on a six-point Likert-type scale (*completely different vs. completely equivalent*). The overall equivalence was good and highly comparable across job conditions.

Procedure and design. Recruiters received an email with a url and personal code that asked for participation in a study on the development of a tool aimed to train/assess recruiters' competencies. To mask the study purpose and to reduce potential item priming, we also included several filler tasks (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). After completing the consent form, participants completed a "work sample test," consisting of the *resumé-sifting task* (experimental task) and three

¹More detailed descriptions are available from the first author on request.

TABLE 1
Descriptives, Reliabilities, and Intercorrelations Among Study Variables (Study 1)

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Dutch-female suitability	3.44	.56	.88											
2. Arab-female suitability	3.15	.73	-.08	.90										
3. Dutch-male suitability	3.32	.56	.20	-.37**	.83									
4. Arab-male suitability	2.90	.65	-.46**	.41**	-.03	.90								
5. Explicit ethnic prejudice	2.70	1.20	.18	-.44**	.29**	-.35**	.90							
6. Explicit sexism	2.90	1.29	-.13	.29*	-.01	.42**	.03	.71						
7. Arab-IAT ^a	<i> .60 </i>	.41	.12	-.10	-.04	-.20	-.07	-.14	.82					
8. Gender-roles IAT ^b	<i> .47 </i>	.38	.22	-.05	-.19	-.25	-.12	-.15	.36**	.78				
9. Social desirability	3.18	.60	-.10	-.24	.12	-.06	.25	.03	.00	.05	.70			
10. Age of recruiter ^b	3.53	1.81	-.10	-.06	-.02	.02	.09	.16	-.14	-.30*	-.05	–		
11. Sex of recruiter ^c	1.58	.49	.20	-.14	-.04	-.28*	.05	-.69**	.03	-.01	-.09	-.25	–	
12. Work experience ^d	2.09	.44	.15	.19	.11	.08	.25	.35**	-.02	.13	.23	-.16	-.16	–

Note. Arab-IAT = implicit prejudice against Arabs; gender-roles IAT = implicit sexism (i.e., implicit beliefs on gender roles). Cronbach's alphas are presented in italics along the diagonal.

^aD600-scores (see for details: De Houwer & De Bruycker, 2007; Greenwald, Nosek, & Banaji, 2003).

^bAge: 1 = 20–25, 2 = 26–30, 3 = 31–35, and so on until 9 = more than 65 (in 5-year increments through official retirement age).

^cSex: 1 = Male; 2 = Female;

^dWork experience: 1 = 0–3 years, 2 = 3–10 years, 3 = more than 10 years of experience with recruiting.

* $p < .05$. ** $p < .01$.

filler tasks. The resumé-sifting task consisted of a [2 (Ethnicity) \times 2 (Sex)] \times 2 (Client contact) mixed-factor design. Ethnicity (Dutch/Arab) and sex (Male/Female) were manipulated within-subjects via names and affiliations on resumé; client contact was manipulated between-subjects via job descriptions in advertisements. Recruiters were randomly assigned to one of the contact conditions and asked to evaluate four resúmes (with ethnicity and sex crossed) by rating applicants' job suitability.

After the resumé-sifting task, recruiters completed three filler tasks, namely (a) a knowledge test about the ethical code of conduct in recruitment, (b) an evaluation of (other) job advertisements, and (c) an evaluation of resumé elements (e.g., how important resumé length and font is to recruiters). At the end of the third filler task, we asked recruiters to indicate applicants' personal background information, such as sex (*manipulation check*), educational level, and ethnicity (*manipulation check*) from resúmes.

After the filler tasks and the manipulation check, recruiters were asked to fill-out an "opinion survey" that included the explicit prejudice measures and the social desirability items among other filler items (e.g., on training motivation). Having recruiters complete the experimental task (i.e., resumé sifting) before the fillers and the explicit prejudice measures was done to avoid any priming effect on our key measure of interest. Finally, recruiters completed the demographic measures. Two to 4 weeks later we presented the recruiters with the implicit prejudice measures (IATs). We could not collect the IATs over the web (i.e., as we used the E-prime program), therefore, we made an appointment with each recruiter to collect the implicit prejudice data at the recruiters' offices (individually). We ended the study with an open-ended probe to ask for any suspicion regarding study purpose.

Measures. Upon completion of the work sample test, recruiters responded to questions on the following measures, using a 1 = *strongly disagree* to 5 = *strongly agree*) response scale unless otherwise noted.

Job suitability. A four-item measure of job suitability from Derous et al. (2009) with a five-point response scale (1 = *not suitable at all* to 5 = *very suitable*) was used. An example item is "Given all the information you read about this applicant, how suitable do you believe this applicant is for this function?" Cronbach's alpha across the four resumé versions ranged from .83 to .90.

Explicit ethnic prejudice. Explicit ethnic prejudice toward Arabs was measured using four items taken from the Modern Racism Scale (McConahay, Hardee, & Batts, 1981) and adapted for Arab descent (MRS-A; Derous et al., 2009). An example item is "Discrimination against Arab minorities is no longer a problem in the Netherlands." Cronbach's alpha was .90.

Explicit sexism. Explicit sexism was assessed with the Modern Sexism Scale of Swim, Aikin, Hall, and Hunter (1995). An example item is “The Dutch society has reached the point where women and men have equal opportunities for achievement.” Cronbach’s alpha in this sample was .71.

Social desirability. Social desirability was included as a control variable and assessed with the Impression Management Scale of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991). An example item is “When I hear people talking privately, I avoid listening.” Cronbach’s alpha was .70.

Following the nested model comparison methodology (e.g., Brown, 2005), a series of confirmatory factor analyses (LISREL v. 9.1.) using the maximum likelihood estimation method showed that respondents were able to distinguish between the various scales they completed during Study 1. Results showed that the basic, four-factor model in which items only loaded on their respective and intended underlying factor (i.e., Job Suitability, Explicit Ethnic Prejudice, Explicit Sexism, and Social Desirability) had an acceptable fit, $\chi^2(242) = 387.93, p = .00$, root mean square error of approximation RMSEA = .09, comparative fit index CFI = .89. The four-factor model fit better than a three-factor model in which job suitability items loaded on one factor, ethnic prejudice and sexism items loaded on a second factor, and social desirability items loaded on a third factor, $\chi^2(245) = 540.33, p = .00$, RMSEA = .14, CFI = .77 ($\Delta\chi^2 = 152.40, \Delta df = 3, p = .00$). The four-factor model also had a significantly better fit compared to a one-factor model in which all items loaded on one single latent factor, $\chi^2(248) = 684.16, p = .00$, RMSEA = .17, CFI = .67 ($\Delta\chi^2 = 296.23, \Delta df = 6, p = .00$).

Demographics. Demographics included recruiters’ age, sex (1 = *male*; 2 = *female*), ethnicity (1 = *Dutch*; 2 = *non-Dutch*), and work experience. Age was measured on a nine-point Likert-type scale (1 = 20–25 to 9 = *more than 65*). Work experience was measured with one item (i.e., “As a recruiter, how much experience do you have in recruiting applicants?”) on a three-point scale (1 = 0–3 years of experience; 2 = between 3–10 years of experience; 3 = more than 10 years of experience with recruiting).

Implicit prejudice. Several weeks later, implicit prejudice was measured with two IATs (i.e., Arab-IAT for ethnic prejudice against Arabs; gender-roles IAT for sexism) that were modeled after Greenwald, Nosek, and Banaji (2003) and presented in counterbalanced order. The IAT measures the strength of automatic associations respondents hold between a concept (e.g., Arab names/male words) and an attribute (e.g., pleasant words/career-related words). Implicit prejudice is concluded when response times are shorter in the compatible pairing condition (e.g., “she” and “family”) when highly associated concepts and attributes are paired and longer in the incompatible condition (e.g., “she” and “office”) when

loosely associated concepts and attributes are paired (see Greenwald et al., 1998, for more information on the IAT measure and procedure).

IATs were conducted in E-prime v2.0. We administered the Arab-IAT that was also used by Derous et al. (2009; Dutch version), and we modeled the gender-roles IAT for sexism after the gender-roles IAT of Rudman and Kilianski (2000). The Arab-IAT paired Arab/Dutch names with pleasant/unpleasant words. Consistent with the lack-of-fit model of Heilman (Heilman, 1983; Heilman & Eagly, 2008), the gender-roles IAT paired female/male words with career-/domestic-related words. Specifically within employment contexts, sex bias in evaluations of women arises from the perceived mismatch between their inferred female attributes and role requirements. Note that all stimuli were pretested to be matched on valence. Following De Houwer and De Bruycker (2007), we calculated a D600 IAT score² for each of our participants using the new scoring algorithm recommended by Greenwald et al. (2003). This improved algorithm is recommended over more conventional methods as it is less contaminated by extraneous variables and it corrects for spuriously extreme IAT scores for slow responders.

Results

Preliminary analyses. Manipulation checks showed that participants perceived the applicants' ethnicity and sex as intended. In addition, we tested the perceived educational level of the applicants as a proxy of job demands (1 = *low*; 4 = *high*), which had a mean of 2.05 ($SD = .18$), indicating that this was perceived as somewhat of a lower-level job.

We also investigated whether the results of both IATs were in the expected direction. The averaged D600 values were .60 for the Arab-IAT and .47 for the gender-roles IAT. A series of *t*-tests showed that both of the D600 scores differed significantly from zero, suggesting that participants had negative implicit attitudes toward Arabs ($t[57] = -11.24$, $p = .00$, Cohen's $d = 1.47$), and that participants held rather traditional gender role attitudes toward women and men ($t[57] = 9.56$, $p = .00$, Cohen's $d = 1.26$).

To calculate reliabilities for each of the two IATs, we followed the procedure by De Houwer and De Bruycker (2007) using an odd-even split-half procedure. For each IAT, we first listed all the trials by order of appearance, separately for each stimulus type (e.g., Arab, Dutch, positive, negative), test block (e.g., Arab-positive, Dutch-positive), and participant. Separate IAT effects (operationalized by the adapted D600 measure) were

²The D600 scores as used by De Houwer and De Bruycker (2007) are equal to the D4 improved algorithm scores as mentioned by Greenwald et al. (2003).

TABLE 2
Mean Job Suitability Scores and Standard Deviations by Experimental Conditions (Applicant Profile \times Client Contact) in Study 1

Applicant profile	Job type			
	Low client contact ($n = 28$)		High client contact ($n = 29$)	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Dutch-female resumé	3.35	.44	3.53	.65
Arab-female resumé	3.21	.65	3.08	.82
Dutch-male resumé	3.32	.42	3.34	.69
Arab-male resumé	3.05	.57	2.78	.66

TABLE 3
Results of Mixed Analyses of Variance for Job Suitability (Study 1)

Source	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Between-subjects</i>				
Client contact (A)	1	.38	.54	.00
Error (A)	55	(.09)		
<i>Within-subjects</i>				
Ethnicity (B)	1	9.90**	.00	.15
A \times B	1	1.97	.17	.03
Error (B)	55	(.71)		
Sex (C)	1	3.95*	.05	.07
A \times C	1	.75	.39	.01
Error (C)	55	(.42)		
B \times C	1	1.74	.19	.03
A \times B \times C	1	.01	.91	.00
Error (B \times C)	55	(.13)		

Note. *df* = degrees of freedom; *F* = *F*-ratio; η^2 = measure of strength of relationship. Values in parentheses represent mean square errors. Additional analyses showed that recruiter sex and work experience had no effect on the study findings.

* $p < .05$. ** $p < .01$.

calculated for odd and even subsets. Subsequently, the split-half reliability was calculated by correlating these two sets of scores and by applying a Spearman–Brown correction. The average split-half reliability in Study 1 was .82 for the Arab-IAT and .78 for the gender-roles IAT.

Main findings. Table 1 presents descriptive statistics and correlations among study variables; Table 2 provides ratings by condition. We found main effects for ethnicity ($F[1, 55] = 9.90$, $p = .00$, $\eta^2 = .15$) and sex ($F[1, 55] = 3.95$, $p = .05$, $\eta^2 = .07$; Table 3). Job suitability ratings were higher for Dutch compared to Arab applicants but lower

for men than women. There was no significant interaction effect of ethnicity and sex, but there was a cumulative effect of both separate main effects that leads to the lowest ratings for Arab men (Table 1). Three- and two-way interactions between ethnicity and sex with client contact were nonsignificant (Hypothesis 2 not supported). However, support was found for Hypothesis 3a ($F[1, 53] = 14.2, p = .00, \eta^2 = .21$), as Arab candidates were rated higher when recruiters' explicit prejudice against Arabs was low. No support was found for Hypothesis 4a: Implicit prejudice toward Arabs did not affect findings ($F[1, 52] = 1.30, p = .26$). Further, no moderating effects were found for explicit sexism ($F[1, 53] = .53, p = .47$; Hypotheses 3b not supported). However, implicit sexism did moderate job suitability ratings, ($F[1, 51] = 5.25, p = .03, \eta^2 = .09$). Contrary to what we expected, women received *higher* job suitability scores when implicit sexism was high (Hypothesis 4b not supported). This two-way interaction, however, was further qualified by a three-way interaction with client contact.

In addition, we explored whether moderating effects of the implicit prejudice measures were stronger than those of the explicit prejudice measures and whether moderating effects depended on job context characteristics (i.e., client contact). Moderating effects were stronger for explicit prejudice against Arabs than for implicit prejudice against Arabs (Hotellings $t[54] = 2.17, p = .03$). The moderating effects of implicit sexism and explicit sexism did not differ significantly from each other (Hotellings $t[54] = -.80, p = .42$). However, a three-way interaction was found between applicants' sex, recruiters' implicit sexism, and the amount of client contact ($F[1, 51] = 7.54, p = .00, \eta^2 = .13$). This three-way interaction showed that the moderating effect depended on the job condition. Specifically, when recruiters had stronger traditional, implicit beliefs regarding gender roles, female applicants were rated higher and men were rated lower when they applied for a high client contact position compared to a low client contact position (Figure 1). That is, the matching principle of considering the stereotype in relation to the job characteristics appears to be operational for those who endorsed traditional gender roles. For explicit prejudice against Arabs, no such three-way interaction was found and there were also no three-way interactions for explicit sexism and implicit prejudice against Arabs.³

³Although the sample size of Study 1 is modest, our primary hypotheses were within-subjects, and both a priori and post hoc power analyses (O'Keefe, 2007) showed that we had sufficient power (above the .80 cutoff mentioned by Cohen, 1992) to detect the hypothesized relations with small-to-moderate effect sizes. To further investigate the robustness of Study 1 findings, we performed the analyses of the basic models in a bootstrap procedure using the R program (R version 3.0.1; Fox & Weisberg, 2011; R Core Team, 2013). Results are the same as for the original analysis.

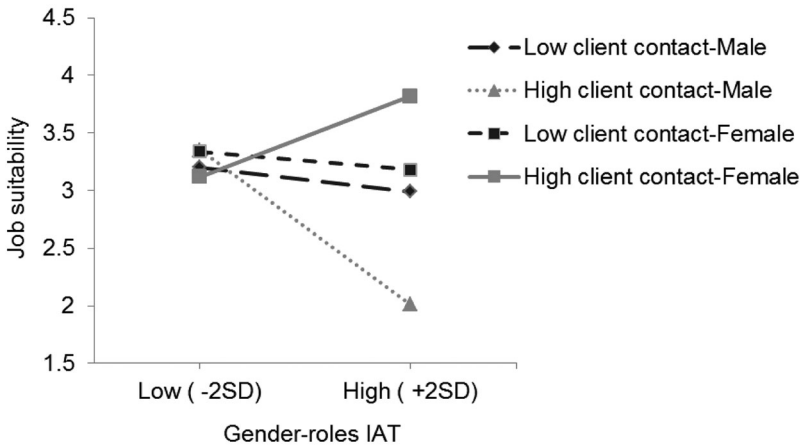


Figure 1: Moderating Effects of Recruiters' Implicit Sexism on Job Suitability Ratings of Male/Female Applicants Applying for Low-Demand Jobs With Low/High Client Contact (Study 1).

Discussion

Evidence was found for bias against Arabs and men, with the cumulative effect pointing in the direction of the subordinate male target hypothesis (i.e., lower job suitability rates for Arab men). Overall, Arabs received significantly lower job suitability scores than their Dutch counterparts in both the high and low client contact conditions. As expected, explicit ethnic prejudice moderated job suitability ratings such that high prejudiced recruiters gave lower job suitability ratings to Arabs than low prejudiced raters. We also found that the matching principle of considering the social category stereotype and the job held only for those high in implicit sexism. That is, women were rated higher by those with greater implicit gender role endorsement for “matching” (i.e., high contact) jobs. This is in line with findings of Rudman and Glick (2001), who showed that one who does not consciously endorse conventional gender stereotypes (e.g., the belief that women are more domestic-oriented and caring whereas men are more career-minded and agentic) may nonetheless act in a prejudiced way because of implicit gender-role stereotypes. Finally, Study 1 showed that there were more effects of explicit prejudice against Arabs than for implicit prejudice against Arabs. The effects of implicit sexism were larger than those of explicit sexism but only so if applicants applied for a high client contact position. This suggests it might be somewhat less acceptable for recruiters to be openly sexist than to be openly racist against Arab minorities.

In sum, Study 1 findings show hiring discrimination against ethnic minority applicants in low-demand jobs, echoing other research (Altonji, 2005; Andriessen et al., 2012; De Beijl, 2000; Derous et al., 2009). Because low-demand jobs do not require high ability and educational credentials, assumptions regarding minority candidates' lower capabilities might have been seen as matching the job. Employment discrimination may also intensively manifest in situations where majority group members are directly or symbolically threatened by the advancement of ethnic minority members, such as in high-demand jobs (Dovidio & Gaertner, 1996; Hosoda et al., 2003).

Thus, one goal of Study 2 was to investigate whether Study 1 findings regarding multiple categories and matching effects occurred for high-demand jobs. Study 2 also investigated two additional factors, ethnicity salience and external motivation to respond without prejudice/sexism, discussed next.

Study 2

Ethnicity Salience

The traditional view of bias considers membership in a particular outgroup as having the same effects on employment outcomes for all members of that outgroup. This traditional view of bias does not take into account *within-category differences* (i.e., differences between members from the same social category). However, research has shown that broad stereotypes can be differentiated into subtypes (Brewer, Dull, & Lui, 1981; Devine & Baker, 1991) when provided with cues or prompts to categorize (e.g., Black athlete, ghetto Black). Category salience (e.g., by the numbers of ethnic cues) should further increase attention to that categorization (Kulik et al., 2007). Thus, when a recruiter is confronted with multiple cues related to an individual's categorization (like ethnic names, pictures, ethnic affiliations), it increases the accessibility of the category and the likelihood that he/she will engage in attending to that category.

Similarly, research has suggested that minority group members may be rejected in proportion to their "outgroupness" (Crisp & Hewstone, 1999; Kaiser & Pratt-Hyatt, 2009; Urban & Miller, 1998). A recent series of studies offers evidence that multiple cues of category membership lead to greater discrimination. For instance, in a lab setting, Uhlmann et al. (2002) showed stronger implicit preferences for light-skinned ethnic minorities over their dark-skinned counterparts. Purkiss, Perrewe, Gillespie, Mayes, and Ferris (2006) found that two ethnic cues (name and accent) led to more negative interviewer reactions than one cue. In both simulated (Deros et al., 2009) and actual hiring contexts (Deros

& Ryan, 2012), recruiters reported more negative attitudes toward ethnic minority applicants identified as such through multiple cues (names and affiliations) compared to their less ethnically-identified counterparts (i.e., with only one cue as to ethnicity).

Recruiters can easily infer undisclosed personal characteristics, such as ethnic identification information, from resumé characteristics like name and affiliations. Names are an important part of personal/social identity and have been evidenced as a direct antecedent of employment discrimination (Bertrand & Mullainathan, 2004; Blommaert et al., 2012). Research has also shown that employers may make negative personnel decisions based on applicants' ethnic group affiliations (Deraus et al., 2009; Dovidio & Gaertner, 2000). Because category salience will affect attention to that category (Kulik et al., 2007) and because majority members may reject minority group members in proportion to their degree of outgroupness (e.g., Kaiser & Pratt-Hyatt, 2009), we expect that:

Hypothesis 5: Job suitability ratings will be lower for resumé of applicants with strong Arab identity cues (i.e., an Arabic name and group affiliations) when compared to those of equally qualified applicants with all Dutch cues (i.e., Dutch name and affiliations) or with mixed Arab and Dutch cues (either Arabic name or affiliations only).

Further, the salience of one category might affect attention to other social categories. In this case, ethnicity salience might affect the extent to which sex is attended to in categorizing individuals. Kulik et al. (2007) proposed that the category with the greatest number of cues will dominate impression formation in a multiple-category situation. Therefore, in Study 2, we further explored Hypothesis 1 regarding multiple categorization effects (i.e., subordinate male target hypothesis). Specifically, we investigated whether the subordinate male target effect would depend on the number of ethnic cues (i.e., ethnicity salience) or whether only a main effect for ethnicity would occur.

External Motivation to Reduce Prejudice

Kulik et al. (2007) suggested motivation to control prejudice affects category activation and inhibition. That is, individuals are sensitive to situational demands like social norms regarding how appropriate it is to express prejudice and will inhibit activation of categories for which there is clear normative pressure not to use them in decision-making (Brief et al., 2000). Plant and Devine (1998) developed the External Motivation to Respond without Prejudice Scale, which refers to perceivers' effortful

cognitive processes to counteract automatically activated stereotypes because of externally situated motives (like peer pressure). For instance, a recruiter might wish to act unprejudiced toward ethnic minority applicants and female applicants because of professional standards, organizational norms, and goals regarding diversity. Similarly, a recruiter might actively seek to avoid prejudice toward ethnic and female applicants to avoid social disapproval of peers.

Few studies, however, have investigated the actual moderating role of external motivation to control for prejudice on *actual* recruiters' discriminatory behavior (see Ziegert & Hanges, 2005 for an example among undergraduates). Furthermore, in a lab study, Derous, Ryan, and Nguyen (2012) found moderating effects of motivation to control for ethnic prejudice, but they did not test this for sexism. Based on the External Motivation to Respond without Prejudice Scale, Klonis, Plant, and Devine (2005) developed the External Motivation to Respond without Sexism Scale, which relates to both sexism and social evaluative concerns (like fear of negative evaluation). In Study 2, we therefore hypothesized moderating effects for recruiters' external motivation to control prejudice and sexism, such that:

Hypothesis 6: Job suitability ratings will be higher for Arab applicants when recruiters' external motivation to respond without prejudice against Arab ethnics is high (Hypothesis 6a) and will be higher for female applicants when external motivation to respond without sexism against women (Hypothesis 6b) is high.

In summary, Study 2 allowed us to examine the same hypotheses and exploratory questions as in Study 1 (Hypotheses 1–4b) but for a job with greater demands. Study 2 also investigated two additional hypotheses (Hypotheses 5–6b) regarding other contextual (ethnic salience) and individual (motivation to control prejudice and sexism) influences on multiple category activation and inhibition.

Method

Participants. The sample consisted of 124 non-Arab/Dutch recruiters (57.8% male), identified and stimulated to participate in the same way as in Study 1. All participants were experienced in recruiting and had between 3 and 10 years of work experience as a recruiter (Table 4).

Procedure, design, and measures. Study 2 used a [3 (Ethnicity) \times 2 (Sex)] \times 3 (Client contact) mixed-factor design. Ethnicity (Dutch/Mixed Arab–Dutch/Arab) and sex (male/female) were measured within subjects, and client contact (low/moderate/high) was manipulated between subjects.

TABLE 4
Descriptives, Reliabilities, and Intercorrelations Among Study Variables (Study 2)

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1. Dutch-female suitability	3.85	.74	.90																
2. Mixed (Arab/Dutch)-female suitability	3.74	.60	.35**	.87															
3. Arab-female suitability	3.76	.82	.20*	.46**	.89														
4. Dutch-male suitability	3.64	.87	.01	.09	.06	.91													
5. Mixed (Arab/Dutch)-male suitability	3.61	.58	.16	.26**	.28**	.20*	.85												
6. Arab-male suitability	3.65	.77	.32**	.38**	.40**	-.16	.23*	.83											
7. Explicit ethnic prejudice	2.31	.81	.08	-.07	-.09	.00	-.20*	-.14	.78										
8. EMS	2.24	.77	.07	.10	-.02	.11	-.23**	-.25**	.17*	.83									
9. Explicit sexism	2.94	.83	.02	-.16	-.08	-.10	-.14	.01	.15	-.19*	.89								
10. EMS-S	1.92	.66	-.09	-.01	-.14	.06	-.23*	-.34**	.31**	.66**	-.11	.86							
11. Arab-IAT ^a	.40	.34	.04	.14	.06	-.06	-.00	-.14	-.19*	-.05	-.12	-.01	.77						

(continued)

TABLE 4 (continued)

	<i>Mean</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
12. Gender-roles IAT ^a	1.38	.37	-.13	.14	.08	-.10	.03	.04	-.09	-.04	-.11	-.02	.18*	.82				
13. Social desirability	2.98	.59	.02	-.18	.05	-.10	.07	.01	.07	-.12	.21*	-.05	.03	-.12	.67			
14. Age of recruiter ^b	4.3	2.09	-.09	-.17	-.20*	.08	-.10	.01	-.05	-.02	-.02	.08	.00	-.05	.21*	-		
15. Sex of recruiter ^c	1.42	.50	-.10	-.09	-.04	-.03	.07	-.04	-.06	.03	-.13	-.10	.08	.13	.06	-.08	-	
16. Work experience ^d	2.16	.54	.05	.10	-.05	.06	.03	.08	.06	-.12	.14	-.14	.15	-.09	-.02	.21*	.02	-

Note. EMS = External motivation to respond without prejudice; EMS-S = external motivation to respond without sexism; Arab-IAT = implicit prejudice against Arabs; gender-roles IAT = implicit sexism (i.e., implicit beliefs on gender roles). Cronbach's alphas are presented in italics on the diagonal.

^aD600-scores (see for details: De Houwer & De Bruycker, 2007; Greenwald, Nosek, & Banaji, 2003).

^bAge: 1 = 20–25, 2 = 26–30, 3 = 31–35 and so on until 9 = more than 65 (in 5-year increments through official retirement age).

^cSex: 1 = Male; 2 = Female.

^dWork experience: 1 = 0–3 years, 2 = 3–10 years, 3 = more than 10 years of experience with recruiting.
 $p < .05$. ** $p < .01$.

Whereas Study 1 presented two low-demand jobs, Study 2 presented three jobs that were higher in demands but that differed in amount of external client contact. Study 2 presented one consultancy job with little external client contact, one consultancy job with moderate client contact, and one consultancy job with high levels of external client contact. The procedure was similar to Study 1. We provided a distinct study purpose to the recruiters before we started. To further mask the study purpose and to reduce potential item priming, we also included several filler tasks and we administered all measures after the resumé screening task.

The measures were the same as in Study 1, with the addition of measures of Motivation to Respond without Prejudice and without Sexism (from Plant & Devine, 1998, and Klonis et al., 2005, respectively). The External Motivation to Respond without Prejudice (EMS) Scale is a five-item scale with a 1 = *disagree* to 5 = *agree* response format; an example item is, "Because of today's politically correct standards, I try to appear non-prejudiced toward Arab minorities." The External Motivation to Respond without Sexism Scale (EMS-S) has a similar format with 10 items. An example item is, "Because of today's politically correct standards, I try to appear nonsexist toward women." Cronbach's alpha was .83 for EMS and .86 for EMS-S (see Table 4 for reliabilities of study variables).

As in Study 1, a series of confirmatory factor analyses (LISREL v. 9.1) with maximum likelihood estimation showed that respondents were able to distinguish between the various scales they completed during Study 2. The basic, six-factor model in which items only loaded on their respective and intended underlying factor (i.e., Job Suitability, Explicit Ethnic Prejudice, External Motivation to Respond Without Prejudice, Explicit Sexism, External Motivation to Respond Without Sexism, and Social Desirability) had an acceptable fit ($\chi^2[667] = 1252.50, p = .00, RMSEA = .08, CFI = .89$). The six-factor model had a significantly better fit compared to a three-factor model in which job suitability items loaded on one factor, ethnic prejudice and sexism items were allowed to load on a second factor, and social desirability items loaded on a third factor ($\chi^2[677] = 1764.40, p = .00, RMSEA = .11, CFI = .79; \Delta\chi^2 = 511.90, \Delta df = 10, p = .00$). The basic, six-factor model also had a better fit compared to a one-factor model in which all items loaded on one single latent factor ($\chi^2[680] = 1973.56, p = .00, RMSEA = .12, CFI = .75; \Delta\chi^2 = 209.16, \Delta df = 13, p = .00$; Brown, 2005).

Results

Preliminary analyses. Manipulation checks were successful for applicants' inferred ethnicity and sex from resumé information. Applicants' perceived educational level was also in line with the intended educational level ($M = 3.02, SD = .12; 1 = low, 4 = high$). The averaged D600 values

TABLE 5
Mean Job Suitability Scores and Standard Deviations by Experimental Conditions (Applicant Profile × Client Contact) in Study 2

Applicant profile	Job type					
	Low client contact (<i>n</i> = 38)		Medium client contact (<i>n</i> = 44)		High client contact (<i>n</i> = 40)	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Dutch-female resumé	3.87	.78	4.06	.63	3.62	.76
Mixed (Dutch/Arab)-female resumé	3.81	.59	3.91	.58	3.46	.56
Arab-female resumé	3.90	.68	4.00	.80	3.36	.83
Dutch-male resumé	3.97	.69	3.07	.82	3.96	.75
Mixed (Dutch/Arab)-male resumé	3.83	.54	3.43	.60	3.59	.55
Arab-male resumé	3.80	.67	3.97	.76	3.16	.63

were .40 for the Arab-IAT and .38 for the gender-roles IAT. As in Study 1, recruiters' implicit attitudes toward Arabs were negative ($t[121] = -13.04, p = .00$, with Cohen's $d = 1.18$). Participants also held rather traditional gender role attitudes toward women and men ($t[121] = -.11.25, p = .00$, with Cohen's $d = 1.01$). The split-half reliability estimates for the Arab-IAT and the gender-roles IAT were .77 and .82, respectively.

Main findings. Table 4 presents descriptive statistics and correlations among all study variables, and Table 5 presents means by condition. Table 6 indicates there was no effect for ethnicity ($F[2, 238] = 1.20, p = .30$), but there was a main effect for sex ($F[1, 119] = 10.54, p = .00; \eta^2 = .08$), with women rated higher than men. Therefore, Hypothesis 1 and Hypothesis 5 were not supported. However, we found significant two- and three-way interactions for ethnicity, sex, and client contact (Table 6). First, there was a significant two-way interaction of ethnicity with client contact ($F[4, 238] = 12.68, p = .00, \eta^2 = .18$). Contrast tests showed that—compared to all Dutch cue profiles ($F[2, 119] = 18.78, p = .00, \eta^2 = .24$), and the mixed Arab–Dutch cue profiles ($F[2, 119] = 12.18, p = .00, \eta^2 = .17$)—high ethnically salient applicants received lower ratings when external client contact was high than when client contact was low or moderate. Second, a significant two-way interaction of client contact with sex ($F[2, 119] = 21.06, p = .00, \eta^2 = .26$) indicated that when client contact was moderate, women were rated significantly higher than male applicants, but not when client contact was low (equal ratings for men/women) or high (slight preference for women). However, when

TABLE 6
Results of Mixed Analyses of Variance for Job Suitability (Study 2)

Source	<i>df</i>	<i>F</i>	<i>p</i>	η^2
<i>Between subjects</i>				
Client contact (A)	2	1.16**	.00	.11
Error	119	(.16)		
<i>Within subjects</i>				
Ethnicity (B)	2	1.20	.30	.01
A × B	4	12.68**	.00	.18
Error (B)	238	(.38)		
Sex (C)	1	10.54**	.00	.08
A × C	2	21.06**	.00	.26
Error (C)	119	(.31)		
B × C	2	.25	.78	.00
A × B × C	4	8.24**	.00	.12
Error (B × C)	238	(.39)		

Note. *df* = degrees of freedom; *F* = *F*-ratio; η^2 = measure of strength of relationship. Values in parentheses represent mean square errors. Additional analyses showed these findings were not affected by recruiter sex and work experience.

* $p < .05$. ** $p < .01$.

considering the three-way interaction of ethnicity, sex, and client contact, support was found for Hypothesis 2 ($F[4, 238] = 8.24, p = .00, \eta^2 = .12$), such that in the high client contact condition, the difference in ratings between female and male Arabs was higher than in the low client contact conditions (Figure 2). In other words, the proposed effects of multiple categories (Hypothesis 1) and category salience (Hypothesis 5) are affected by job characteristics.

Supporting Hypotheses 3a and 6a, explicitly-measured ethnic prejudice against Arabs ($F[2, 112] = 4.28, p = .02, \eta^2 = .07$) and external motivation to respond without prejudice ($F[2, 114] = 3.47, p = .03, \eta^2 = .06$) moderated ethnicity such that Arab applicants were rated higher when explicit prejudice was low and when motivation to respond without prejudice was high. Finally, implicit ethnic prejudice and sexist attitudes (both implicitly measured and explicitly measured) did not affect findings. Therefore, Hypotheses 3b, 4, and 6b were not supported.⁴

As in Study 1, we explored whether the moderating effect of the implicit prejudice measures were stronger than those of the explicit prejudice measures and whether any moderating effect of prejudice depended on the job context (i.e., degree of client contact). Results showed no significant moderating effect for implicitly-measured prejudice. Furthermore, and

⁴Study 2 power was sufficient to detect expected effects (O'Keefe, 2007). Results of bootstrap analyses also supported robustness of Study 2 findings.

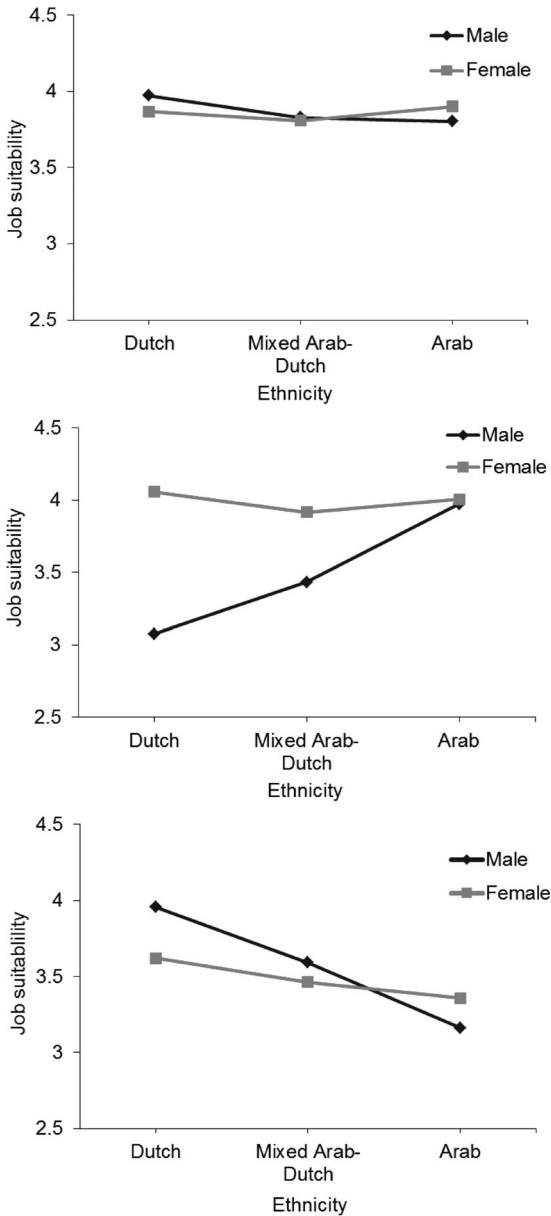


Figure 2: Interaction of Ethnicity (Dutch/Mixed Arab–Dutch/Arab) and Sex (Male/Female) by Client Contact (Low/Moderate/High; Study 2).

Note. The first graph (top) represents effects when client contact is low; the second graph (middle) represents effects when client contact is moderate; the third graph (bottom) represents effects when client contact is high.

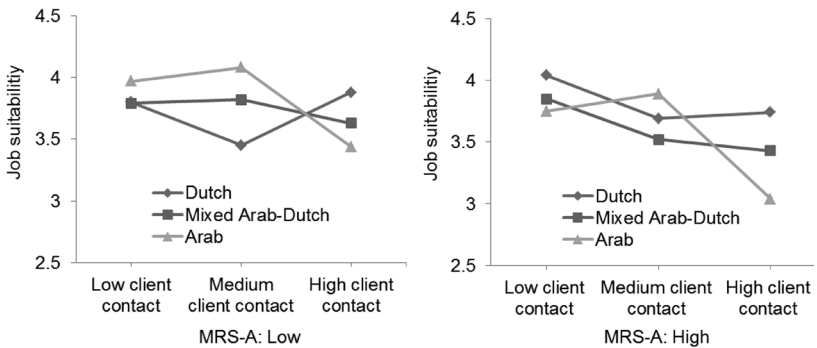


Figure 3: Job Suitability Ratings of Dutch, Mixed Arab–Dutch, and Arab Applicants Applying for High-Demand Jobs with Low/Medium/High Client Contact.

Note. The left graph represents these effects when explicit ethnic prejudice (MRS-A) is low (-2 SD MRS-A); the right graph represents these effects when explicit ethnic prejudice is high ($+2$ SD MRS-A; Study 2).

as in Study 1, no significant three-way interaction was found between applicants' ethnicity, recruiters' explicit prejudice against Arabs, and client contact ($F[4, 226] = 1.24, p = .29$). The two-way interaction of ethnicity with explicit prejudice against Arabs did not differ significantly between the high and low client contact positions ($F[2, 70] = 2.25, p = .11$; Figure 3). However, significant differences were found between the medium client contact condition and the low client condition on the one hand ($F[2, 76] = 4.53, p = .01, \eta^2 = .11$), and between the medium client contact condition and the high client condition on the other hand ($F[2, 77] = 3.14, p = .04, \eta^2 = .08$). In the medium client contact condition, profiles with more cues to Arab ethnicity were rated higher than mixed Arab–Dutch and Dutch profiles. In the high and low client contact condition, profiles with more cues to Arab ethnicity were rated lower than the others, particularly when explicit prejudice was high. The other three-way interactions were not significant.

Discussion

Certain contexts can exert a stronger biasing influence on social judgments than others (Kulik et al., 2007). As the jobs in Study 1 and Study 2 differed, the findings also were slightly different. However, there were a number of key similarities. First, interactions with client contact levels found in both studies demonstrated that multiple category activation and inhibition might depend on the particular job the applicant applies for and recruiters' prejudice levels (Kulik et al., 2007). Second, in both studies,

moderating effects were found for explicitly-measured prejudice against Arabs. In Study 2, when explicit prejudice toward Arabs was high, Arab applicants received the lowest job suitability ratings in the high and low client contact conditions relative to the moderate client contact condition (Figure 3). This finding may reflect attributional ambiguity. Every resumé included two previous work experiences with one in a front-office environment and one in a back-office environment. In the medium client contact condition, therefore, applicants' job qualifications (i.e., mix of front- and back-office experiences) may have been viewed as a better match to the job requirements (medium client contact) than those in the low and high client contact conditions. Contrary to the high and low client contact job conditions, there is less of a justification for prejudice in the moderate client contact condition as job qualifications may have been seen as fitting the job to a greater extent. Thus, prejudiced raters may consider applicants' more ambiguous job qualifications as a justification for their discrimination in the high and low client contact conditions. Study 2 also added beyond Study 1 in showing a role for motivation to respond without prejudice. Recruiters high in motivation to respond without prejudice gave more positive evaluations to Arab applicants. Whether stereotypes affected hireability ratings not only depended on situational characteristics but also on recruiter individual differences. However, no moderating effect was found for external motivation to respond without sexism.

Overall Discussion

Considering discrimination in the resumé screening phase is important as this might affect the overall quality of applicants and adverse impact of later selection stages (De Corte, 2011). This study adds to this literature by addressing multiple categorization effects in resumé screening. Our study findings support Kulik et al.'s (2007) propositions regarding the complex nature of hiring discrimination where applicant, recruiter, and context characteristics affect the activation and inhibition of social category information.

Findings and Contributions

We first discuss three key findings and similarities across studies and then discuss unique contributions of each. One key finding is that across both low- and high-demand jobs (Studies 1 and 2) Arab women fared better than their male counterparts in ratings, pointing in the direction of the subordinate male target hypothesis, particularly when considering matching of gender role stereotypes and job characteristics. There may be

presumptions that women are better at client contact and interpersonal exchanges, activating the gender category more so than ethnicity in recruiter review of resumé, as suggested by Kulik et al. (2007).

Second, across both studies, explicitly-measured ethnic prejudice affected evaluations. This is also in line with Kulik et al.'s (2007) proposition that individuals will pay more attention to categories for which they have stronger attitudes. Both studies further show evidence for attributional ambiguity effects.⁵ Justification for prejudiced reactions can rely on more external, business-related factors (like client contact/demands) or on more internal, applicant-related factors (like an ambiguous fit of the applicants' job qualifications to the job requirements). Our findings showed evidence for both.

Further, we were able to examine both recruiters' explicit and implicit attitudes. However, across both studies, results do not suggest that unconscious attitudes play a strong role. Effects for unconscious prejudice against Arabs were not found in either study, and in Study 1, unconscious sexism's effect on rating depended on the job context. Across both studies, implicit prejudice for ethnicity and gender role endorsement was relatively large. Even with these relatively large levels of implicit bias, effects were only found for gender-role endorsement and the ratings of women in low-demand jobs. Similarly, in the literature, mixed findings have been reported on the effects of implicit prejudice on hiring decisions, with some studies showing positive effects (Rooth, 2010) and others showing no or mixed findings (Blommaert et al., 2012; Derosus et al., 2009). One explanation for mixed findings may be in the nature of the resumé screening task (i.e., providing job suitability ratings), which—for a large part—is under recruiters' conscious control, something that might have been heightened given recruiter awareness that they were in a research study. Specifically, implicit measures may be better in predicting automatically evoked and nondeliberate forms of discrimination that are less under conscious control (Dovidio, Kawakami, & Gaertner, 2002).

A major difference between the studies was in level of job demands. In low-demand jobs (Study 1), there was a cumulative effect of ethnicity and sex, such that ethnic minority men were rated the lowest, regardless of the client contact level of the job. In high-demand jobs (Study 2),

⁵The paper follows the original description of Snyder, Kleck, Strenta, and Mentzer (1979). A slightly different approach has been presented by Crocker and Major (1989) who applied attributional-ambiguity to the stigmatized instead of the stigmatizing person. They specifically suggested attributional ambiguity as a defense mechanism of the stigmatized to deal with (perceived) prejudice and discrimination in order to protect the self.

interactions of applicants' ethnicity and sex with amount of client contact were found, showing that the factors proposed by Kulik et al. (2007) as affecting multiple category inhibition and activation (i.e., category salience, job matching) do operate in as complex a manner as they suggest. For example, the more cues to ethnicity on the resumé, the greater the hiring discrimination, but only when the amount of client contact was also high. Applicants whose ethnicity is very salient may be perceived as a threat to the legitimacy of the status quo (Shelton & Richeson, 2005), whereas lowering ethnic salience may reduce perceptions of challenge (Jussim, Fleming, Coleman, & Kohberger, 1996). This particularly pertained to the male minority applicant, perhaps because of the assumptions regarding the greater skill level and service-related orientation that might be associated with these jobs. Recruiters appeared to consider matching in terms of stereotypes regarding both the amount of client contact and job demand.

One other difference is that we only examined motivation to control prejudice for high-demand jobs (Study 2), but we did find it played a role in evaluations of Arab applicants. Kulik et al. (2007) proposed that the category most associated with motivation to avoid prejudice (in this case, ethnicity and not sex) should be the one most inhibited in impression formation.

In sum, this study contributes to the literature by expanding studies of discrimination in resumé screening to consider multiple categorization effects and moderators of such effects. Our findings are more in line with the subordinate male target hypothesis than the double jeopardy hypothesis, but go beyond it in pointing out the need to qualify categorization effects with considerations of job characteristics (client contact, job demands) and rater attitudes (ethnic prejudice, sexism).

Limitations and Further Research Opportunities

One limitation of this study is that the specific stereotype held of a particular group or intersection of memberships (e.g., Arab and African American men' stereotype as aggressive, but not Asian men) will affect findings, and we only investigated one ethnic minority group and one other social category (sex). Therefore, cross-validating findings in different marginalized groups is recommended. Whereas the intersectional effects of multiple dimensions of diversity have long been recognized by sociologists and social psychologists (e.g., Sidanius & Veniegas, 2000; Urban & Miller, 1998), in personnel selection, dimensions of diversity have mainly been studied in isolation. The interactive effects of ethnicity with other applicant characteristics (like age) on hiring decisions should be further investigated.

Second, whereas we tested several of Kulik et al.'s premises in this paper, other constructs could be at play. For instance, recruiters' other-group orientation might explain why those who are low in explicit prejudice might rate Arab (men) higher than their respective, highly prejudicial counterparts. Specifically, people who are high in other-group orientation have more favorable attitudes toward interactions with different others (Avery, 2003; Phinney, 1992). In addition, we did not measure recruiters' internal motivation to respond without prejudice, which would tap into a desire to avoid rating bias to maintain a view of the self as nonprejudiced through a dissonance-reduction mechanism. Both controlled and automatic bias can be influenced by factors like the configuration of stimulus cues (e.g., intersection of applicant characteristics) in combination with recruiters' self- and social motives as well as other recruiter characteristics (e.g., recruiter ethnicity). Future research, therefore, may further consider recruiters' motives and demographic status.

Third, another factor that might influence recruiters' information processing is situational constraints (e.g., Blair, 2002; Lowery, Hardin, & Sinclair, 2001; Wittenbrink, Judd, & Park, 2001) like time pressure. Resumes are often screened under time constraints, which may make one more vulnerable to automatic bias, as it depletes one's executive resources (Bertrand, Chugh, & Mullainathan, 2005; Macrae, Milne, & Bodenhausen, 1994). For instance, in a resume screening task (Bertrand et al., 2005), participants selected fewer resumes from African-American applicants, compared to equally qualified White-American applicants, when they felt "in a rush." When one is under pressure, stereotypes may act as heuristics to free up cognitive resources. Therefore, research could further investigate whether activation of categories also depends on situational constraints like recruiters' time constraints.

Finally, our design merits some attention. First, the use of within-subjects designs may be a more realistic approach than the use of between-subjects designs in resume screening settings (Landy, 2008). However, a within-subjects design may have made social category contrasts across resumes more salient to raters. This in turn might have affected (i.e., inflated) hiring discrimination. Second, recruiters were asked to rate suitability of the resumes but were not asked to rank or to make a choice. However, different types of rating and decision-making tasks might result in different levels of bias (Breugh & Starke, 2000). For instance, a top-down standard could require decision-makers to attend more acutely to resume qualifications relative to a minimum competency model, thereby exposing raters to more individuating information. Further, recruiting real recruiters to participate in research tasks is a challenge. Although we recognize our modest sample sizes, we did have sufficient power.

Nevertheless, we recommend future research use larger samples, if possible, to further investigate these findings.

Practical Relevance, Implications, and Conclusion

To avert hiring discrimination, there is a need to move beyond prevalence studies and investigate determinants of discrimination. Resumé screening may be particularly vulnerable to biases because of the limited amount of personalized information about candidates. When more information about the candidate becomes available, category-based biases might have less of a chance to color decisions (Brewer & Harasty Feinstein, 1999; Fiske et al., 1999; Landy, 2008). Increasing categorical complexity by considering multiple criteria for categorization has been offered as a useful strategy to reduce intergroup bias (Hall & Crisp, 2005). Therefore, one could train recruiters to actively generate and use alternative ways to categorize and to evaluate applicants (e.g., experience levels, sectors of employment, number of extracurricular activities), independent from the stigmatizing group(s) they belong to, so that those categories have increased salience.

The use of structured, competency-based resumé screening may be one way to accomplish this. That is, organizations could develop structured screening procedures to lessen social category cue salience (Kulik et al., 2007) by means of competency and experience-based checklists. Efforts to educate organizational decision-makers regarding the value of structured interviews have had some payoff in changing practice (Chapman & Zweig, 2005; Dipboye & Johnson, 2013; Levashina, Hartwell, Morgeson, & Campion, 2014; Van Iddekinge, Raymark, & Roth, 2005). Similar efforts to educate recruiters regarding the value of structured resumé review may help reduce the potential for discriminatory outcomes at this stage and will likely also increase the validity of the process. Therefore, a first recommendation would be the use of structured, competency-based screening processes as part of applicant tracking systems to reduce recruiters' attention to ethnic/gender identifiers (or any other stigmatizing information) in resúmes.

However, our study findings also show the complex nature of hiring discrimination because of interactive effects of applicant characteristics and job characteristics. Organizations should not assume that all recruiters will approach a given job and given subgroup of applicants in a consistent manner, even if the same individuals may evidence no discrimination with other jobs and other subgroups. As illustrated, ethnic minorities might be treated in a different way depending on ethnic salience, job complexity, and the amount of client contact. Another practical implication is for organizations to better consider the likelihood of multiple category effects

as contingent on job context. Direct discussion of job stereotypes and the way ethnic minorities are perceived to match job requirements may be important in some contexts. A second practical recommendation, therefore, concerns recruiter awareness training on job stereotypes, particularly for job types for which recruiters regularly screen minority candidates.

Third, whereas technology may streamline the early stages of hiring processes in a number of ways, many organizations still employ human judges for resumé review. Because recruiters' prejudiced attitudes also matter, a third recommendation would be to screen recruiters on their propensity to discriminate. However, this may be difficult to do, as those who screen resúmes may be in that role because of technical job expertise and/or hiring authority. Identifying recruiters' prejudiced attitudes, however, can still be important in increasing recruiters' awareness of their biases. As our results showed that explicit prejudice came into play in certain contexts (like in high-demand/high client contact jobs or when applicants have rather ambiguous job qualifications), making recruiters aware of when they may be especially vulnerable to biases may be fruitful. Kulik et al. (2007) noted that broader diversity training may also aid in helping individuals recognize the effects of stereotyping. Diversity training, however, could be more effective for recruiters who are more sensitive to social norms regarding how appropriate it is to express prejudice. Therefore, as another recommendation, recruiters could be screened on their motivation to respond without prejudice, even before entering any training program. Perhaps it could also be useful to alert recruiters of their implicit stereotypes that might affect hireability ratings in certain job contexts. However, more research on the role of implicit attitudes in resumé screening is needed before clear practical recommendations can be offered.

Finally, organizations should develop policies and procedures to avert any discriminatory screening practices. Coupled with efforts at increasing recruiter awareness and training, holding resumé screeners accountable for their decisions may also make a difference in reducing bias in resumé screening. For instance, instead of one recruiter, a team of recruiters could be assigned the same piles of resúmes to screen and to discuss. Research on job interviews also showed that the reliability of judgments increased when panel interviews were used instead of one-to-one interviews (e.g., Conway, Jako, & Goodman, 1995). This effect may even be strengthened when demographically diverse recruitment panels are used to screen resúmes.

In conclusion, our study findings shed light on multiple categorization effects and represent some of the first tests of Kulik et al.'s (2007) propositions in the context of resumé screening. Additional research and preventative measures are warranted to protect individuals from

potential illegal and unfair discrimination at very early stages of the screening process.

REFERENCES

- Altonji JG. (2005). Employer learning, statistical discrimination and occupational attainment, Discussion paper No. 3. New Haven, CT: Yale University. Retrieved from <http://www.econ.yale.edu/ddp/ddp000003.htm>
- Andriessen I, Nievers E, Dagevos J. (2012). *Op achterstand: Discriminatie van niet-westerse migranten op de arbeidsmarkt* [Be behind: Discrimination of non-Western immigrants on the Dutch labor market]. The Hague, the Netherlands: Sociaal en Cultureel Planbureau.
- Andriessen I, Nievers E, Faulk LJ, Dagevos J. (2010). *Liever Mark dan Mohammed?* [Better selecting Mark than Mohammed?]. The Hague, the Netherlands: Sociaal en Cultureel Planbureau.
- Ashkanasy NM, Härtel CEJ, Daus CS. (2002). Diversity and emotion: The new frontiers in organizational behavior research. *Journal of Management*, 28, 307–338. doi: 10.1177/014920630202800304
- Avery DR. (2003). Reactions to diversity in recruitment advertising: Are differences black and white? *Journal of Applied Psychology*, 58, 672–679. doi: 10.1037/0021-9010.88.4.672
- Bendick M Jr, Jackson C, Reinoso V. (1994). Measuring employment discrimination through controlled experiments. *Review of Black Political Economy*, 23, 25–48. doi: 10.1007/BF02895739
- Bendick M Jr, Jackson C, Reinoso V, Hodges L. (1991). Discrimination against Latino job applicants: A controlled experiment. *Human Resource Management*, 30, 469–484. doi: 10.1002/hrm.3930300404
- Berdahl J, Moore C. (2006). Workplace harassment: Double jeopardy for minority women. *Journal of Applied Psychology*, 91, 426–436. doi: 10.1037/0021-9010.91.2.426
- Bertrand M, Chugh D, Mullainathan S. (2005). Implicit discrimination. *American Economic Review*, 95, 94–98. doi: 10.1257/000282805774670365
- Bertrand M, Mullainathan S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94, 991–1013. doi: 10.1257/0002828042002561
- Blair IV. (2002). The malleability of automatic stereotypes and prejudice. *Personality & Social Psychology Review*, 6, 242–261. doi: 10.1207/S15327957PSPR0603_8
- Blanton H, Jaccard J. (2006a). Arbitrary metrics in psychology. *American Psychologist*, 61, 27–41. doi: 10.1037/0003-066X.61.1.27
- Blanton H, Jaccard J. (2006b). Arbitrary metrics redux. *American Psychologist*, 61, 62–71. doi: 10.1037/0003-066X.61.1.62
- Blom M, Oudhof J, Bijl RV, Bakker BFM. (Eds.). (2005). *Verdacht van criminaliteit. Allochtonen en autochtonen nader bekeken*. [Suspected from criminal facts: A comparison between ethnic minorities and majorities]. The Hague, the Netherlands: WODC.
- Blommaert L, van Tubergen F, Coenders M. (2012). Implicit and explicit interethnic attitudes and ethnic discrimination in hiring. *Social Science Research*, 41, 61–73. doi: 10.1016/j.ssresearch.2011.09.007
- Booth AL, Leigh A, Varganova E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics*, 74, 547–573. doi: 10.1111/j.1468-0084.2011.00664.x

- Breaugh JA, Starke M. (2000). Research on employee recruitment: So many studies, so many remaining questions. *Journal of Management*, 26, 405–434. doi: 10.1177/014920630002600303
- Brewer MB, Dull V, Lui L. (1981). Perceptions of the elderly: Stereotypes as prototypes. *Journal of Personality and Social Psychology*, 41, 656–670. doi: 10.1037/0022-3514.41.4.656
- Brewer MB, Harasty Feinstein AS. (1999). Dual processes in the cognitive representation of persons and social categories. In Chaiken S, Trope Y (Eds.), *Dual process theories in social psychology* (pp. 255–270). New York, NY: Guilford Press.
- Brief AP, Dietz J, Cohen RR, Pugh SD, Vaslow JB. (2000). Just doing business: Modern racism and obedience to authority as explanations for employment discrimination. *Organizational Behavior and Human Decision Processes*, 81, 72–97. doi:10.1006/obhd.1999.2867
- Brown TA. (2005). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Brunel FF, Tietje BC, Greenwald AG. (2004). Is the Implicit Association Test a valid and valuable measure of implicit consumer social cognition? *Journal of Consumer Psychology*, 14, 385–404. doi: 10.1207/s15327663jcp1404.8
- Chapman DS, Zweig DI. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *PERSONNEL PSYCHOLOGY*, 58, 673–702. doi: 10.1111/j.1744-6570.2005.00516.x
- Cleveland JN, Vescio TK, Barnes-Farrell JL. (2005). Gender discrimination in organizations. In Dipboye RL, Colella A (Eds.), *Discrimination at work: The psychological and organizational bases* (pp.149–176; SIOP Frontiers Series). Mahwah, NJ: Erlbaum.
- Cohen J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Cole MS, Rubin RS, Field HS, Giles WF. (2007). Recruiters' perceptions and use of applicant resumé information: Screening the recent graduate. *Applied Psychology: An International Review*, 56, 319–343. doi:10.1111/j.1464-0597.2007.00288.x
- Conway JM, Jako RA, Goodman DF. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565–579. doi: 10.1037/0021-910.80.5.565
- Crisp RJ, Hewstone M. (1999). Differential evaluation of crossed category groups: Patterns, processes, and reducing intergroup bias. *Group Processes and Intergroup Relations*, 2, 1–27. doi: 10.1177/1368430299024001
- Crocker J, Major B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, 96, 608–630. doi: 10.1037/0033-295X.96.4.608
- Cuddy AJC, Fiske ST, Glick P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, 92, 631–648. doi: 10.1037/0022-3514.92.4.631
- Dagevos J. (2011, March). *De maatschappelijke positie van Marokkaanse Nederlanders anno 2011* [The position of Moroccan-Dutch people in the Netherlands anno 2011]. Paper presented at the SMN Netwerkdag, Utrecht, The Netherlands.
- De Beijl RZ. (2000). *Documenting discrimination against migrant workers in the labor market: A comparative study of four European countries*. Geneva, Switzerland: International Labor Office.
- De Corte W. (2011, June). *An integrated decision-aid for shaping recruitment and selection decisions*. Paper presented at the EAWOP Small Group Meeting on Recruitment and Selection, Athens, Greece.

- De Houwer J, De Bruycker E. (2007). The Implicit Association Test outperforms the extrinsic affective Simon task as an implicit measure of inter-individual differences in attitudes. *British Journal of Social Psychology*, *46*, 401–421. doi: 10.1348/014466606X130346
- de Jong PJ, van den Hout MA, Rietbroek H, Huijding J. (2003). Dissociations between implicit and explicit attitudes toward phobic stimuli. *Cognition and Emotion*, *17*, 521–545. doi:10.1080/02699930302305
- Derous E, Nguyen H-HD, Ryan AM. (2009). Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance*, *22*, 1–24. doi: 10.1080/08959280903120261
- Derous E, Ryan AM. (2012). Documenting the adverse impact of resumé screening: Degree of ethnic identification matters. *International Journal of Selection and Assessment*, *20*, 464–474. doi: 10.1111/ijsa.12009
- Derous E, Ryan AM, Nguyen H-HD. (2012). Multiple categorization in resumé screening: Examining effects on hiring against Arab applicants in field and lab settings. *Journal of Organizational Behavior*, *33*, 544–570. doi 10.1002/job.769
- Devine PG, Baker SM. (1991). Measurement of racial stereotype subtyping. *Personality and Social Psychology Bulletin*, *17*, 44–50. doi: 10.1177/0146167291171007
- Dipboye RL, Johnson SK. (2013). Understanding and improving employee selection interviews. In Geisinger KF (Ed.), *APA handbook of testing and assessment in psychology: Vol 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 479–499). Washington DC: American Psychological Association. doi: 10.1037/14047-027
- Dovidio JF, Gaertner SL. (1996). Affirmative action, unintentional racial biases, and intergroup relations. *Journal of Social Issues*, *52*, 5–75. doi: 10.1111/j.1540-4560.1996.tb01848.x
- Dovidio JF, Gaertner SL. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, *11*, 315–319. doi: 10.1111/1467-9280.00262
- Dovidio JF, Hebl MR. (2005). Discrimination at the level of the individual: Cognitive and affective factors. In Dipboye RL, Colella A (Eds.), *Discrimination at work: The psychological and organizational bases* (pp.11–35; SIOP Frontiers Series). Mahwah, NJ: Erlbaum.
- Dovidio JF, Kawakami K, Gaertner SL. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68. doi: 10.1037//0022-3514.82.1.62
- Esses VM, Jackson LM, Armstrong TL. (1998). Intergroup competition and attitudes toward immigrants and immigration: An instrumental model of group conflict. *Journal of Social Issues*, *54*, 699–724. doi: 10.1111/j.1540-4560.1998.tb01244.x
- Fazio RH, Olson MA. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327. doi: 10.1146/annurev.psych.54.101601.145225
- Feagin JR, Eckberg DL. (1980). Discrimination: Motivation, action, effects, and context. *Annual Review of Sociology*, *6*, 1–20. doi: 0.1146/annurev.so.06.080180.000245
- Fiske ST, Lin M, Neuberg SL. (1999). The continuum model, ten years later. In Chaiken S, Trope Y (Eds.), *Dual process theories in social psychology* (pp. 231–254). New York, NY: Guilford Press.
- Fox J, Weisberg S. (2011). *An R Companion to applied regression* (2nd Edition). Thousand Oaks, CA: Sage.
- Ghumman S, Jackson L. (2010). The downside of religious attire: The Muslim headscarf and expectations of obtaining employment. *Journal of Organizational Behavior*, *31*, 4–21. doi: 10.1002/job.601

- Greenwald AG, McGhee DE, Schwartz JLK. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464
- Greenwald AG, Nosek BA, Banaji MR. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197–216. doi: 10.1037/0022-3514.85.2.197
- Greenwald AG, Poehlman AT, Uhlmann EL, Banaji MR. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17–41. doi: 10.1037/a0015575
- Haines EL, Sumner KE. (2006). Implicit measurement of attitudes, stereotypes, and self-concepts in organizations: Teaching old dogmas new tricks. *Organizational Research Methods*, 9, 536–553. doi: 10.1177/1094428106286540
- Hall NR, Crisp RJ. (2005). Considering multiple criteria for social categorization can reduce intergroup bias. *Personality and Social Psychology Bulletin*, 31, 1435–1444. doi: 10.1177/0146167205276084
- Heilman ME. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior*, 5, 269–298.
- Heilman ME, Eagly AH. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 393–398. doi: 10.1111/j.1754-9434.2008.00072
- Heilman ME, Haynes MC. (2008). Subjectivity in the appraisal process: A facilitator of gender bias in work settings. In Borgida E, Fiske ST (Eds.), *Beyond common sense: Psychological science in the courtroom* (pp. 127–155). Malden, MA: Blackwell.
- Hodson G, Dovidio JF, Gaertner SL. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin*, 28, 460–471. doi: 10.1177/0146167202287004
- Holzer HJ, Ihlantfeldt KR. (1998). Customer discrimination and employment outcomes for minority workers. *Quarterly Journal of Economics*, 113, 835–867. doi:10.1162/0033535398555766
- Hosoda M, Stone DL, Stone-Romero EF. (2003). The interactive effects of race, gender, and job type on job suitability ratings and selection decisions. *Journal of Applied Social Psychology*, 33, 145–178. doi: 10.1111/j.1559-1816.2003.tb02077.x
- Jussim L, Fleming CJ, Coleman LM, Kohberger C. (1996). The nature of stereotypes II: A multiple-process model of evaluations. *Journal of Applied Social Psychology*, 26, 283–312. doi: 10.1111/j.1559-1816.1996.tb01851.x
- Kaiser CR, Pratt-Hyatt JS. (2009). Distributing prejudice unequally: Do Whites direct their prejudice toward strongly identified minorities? *Journal of Personality and Social Psychology*, 96, 432–445. doi: 10.1037/a0012877
- Kihlstrom JF. (2004). Implicit methods in social psychology. In Sansone C, Morf CC, Panter AT (Eds.), *The Sage handbook of methods in social psychology* (pp. 195–212). Thousand Oaks, CA: Sage.
- Klaver J, Mevissen JWM, Odé AWM, Materman S, Weening H. (2005). *Etmische minderheden op de arbeidsmarkt: Beelden en feiten, belemmeringen en oplossingen* [Ethnic minorities in the labor market: Images and facts, obstacles and solutions]. Amsterdam, The Netherlands: Regioplan Beleidsonderzoek.
- Klonis SZ, Plant EA, Devine PG. (2005). Internal and external motivation to respond without sexism. *Personality and Social Psychology Bulletin*, 31, 1237–1249. doi: 10.1177/0146167205275304
- Kulik CT, Roberson L, Perry EL. (2007). The multiple-category problem: Category activation and inhibition in the hiring process. *Academy of Management Review*, 32, 529–548. doi: 10.5465/AMR.2007.24351855

- Landy FJ. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 379–392. doi: 10.1111/j.1754-9434.2008.00071.x
- Levashina J, Hartwell CJ, Morgeson FP, Campion MA. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *PERSONNEL PSYCHOLOGY*, *67*, 241–293. doi: 10.1111/peps.12052
- Lowery BS, Hardin CD, Sinclair S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842–855. doi: 10.1037/0022-3514.81.5.842
- Macrae CN, Bodenhausen GV. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. doi: 10.1146/annurev.psych.51.1.93
- Macrae CN, Milne AB, Bodenhausen GV. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, *66*, 37–47. doi: 10.1037/0022-3514.66.1.37
- McConahay JB, Hardee BB, Batts V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution*, *25*, 563–579. doi: 10.1177/002200278102500401
- Nelson NL, Probst TM. (2004). Multiple minority individuals: Multiplying the risk of workplace harassment and discrimination. In Chin JL (Ed.), *The psychology of prejudice and discrimination: Ethnicity and multiracial identity* (Vol. 2, pp. 193–217). Westport, CT: Praeger/Greenwood.
- Organization for Economic Co-operation and Development. (2008). *Jobs for immigrants (Vol. 2): Labour market integration in Belgium, France, the Netherlands and Portugal*. Paris, France: OECD Publishing. doi: 10.1787/9789264055605-en
- Organization for Economic Co-operation and Development. (2013). *OECD factbook 2013: Economic, environmental and social statistics*. Paris, France: OECD Publishing. doi: 10.1787/factbook-2013-en
- O’Keefe DO. (2007). Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. *Communication Methods and Measures*, *1*, 291–299. doi: 10.1080/19312450701641375
- Paulhus DL. (1991). Measurement and control of response bias. In Robinson JP, Shaver PR, Wrightsman LS (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.
- Phinney JS. (1992). The multigroup ethnic identity measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, *7*, 156–176. doi: 10.1177/074355489272003
- Plant EA, Devine PG. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–832. doi: 10.1037/0022-3514.75.3.811
- Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879–903. doi: 10.1037/0021-9010.88.5.879
- Purkiss SLS, Perrew PL, Gillespie TL, Mayes BT, Ferris GR. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes*, *101*, 152–167. doi: 10.1016/j.obhdp.2006.06.005
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rooth DO. (2010). Automatic associations and discrimination in hiring: Real-world evidence. *Labour Economics*, *17*, 523–534. doi:10.1016/j.labeco.2009.04.005

- Rudman LA. (2008). The validity of the Implicit Association Test is a scientific certainty. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 426–429. doi: 10.1111/j.1754-9434.2008.00081.x
- Rudman LA, Glick P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57, 743–762. doi: 10.1111/0022-4537.00239-4537.00239
- Rudman LA, Kilianski S. (2000). Implicit and explicit attitudes towards female authority. *Personality and Social Psychology Bulletin*, 26, 1315–1328. doi: 10.1177/0146167200263001
- Shelton JN, Richeson JA. (2005). Intergroup contact and pluralistic ignorance. *Journal of Personality and Social Psychology*, 88, 91–107. doi: 10.1111/j.1744-6570.2001.tb00094.x
- Sidanius J, Pratto F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY: Cambridge University Press.
- Sidanius J, Veniegas RC. (2000). Gender and race discrimination: The interactive nature of disadvantage. In Oskamp S (Ed.), *Reducing prejudice and discrimination: The Claremont symposium on applied social psychology* (pp. 47–69). Mahwah, NJ: Erlbaum.
- Snyder ML, Kleck RE, Strenta A, Mentzer SJ. (1979). Avoidance of the handicapped: An attributional ambiguity analysis. *Journal of Personality and Social Psychology*, 37, 2297–2306. doi: 10.1037/0022-3514.37.12.2297
- Son Hing L, Chung-Yan GA, Hamilton LK, Zanna MP. (2008). A two-dimensional model that employs explicit and implicit attitudes to characterize prejudice. *Journal of Personality and Social Psychology*, 94, 971–987. doi: 10.1037/0022-3514.94.6.971
- Swim JK, Aikin KJ, Hall WS, Hunter BA. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, 68, 199–214. doi:10.1037/0022-3514.68.2.199
- Talaska CA, Fiske ST, Chaiken S. (2008). Legitimizing racial discrimination: A meta-analysis of the racial attitude–behavior literature shows that emotions, not beliefs, best predict discrimination. *Social Justice Research: Social Power in Action*, 21, 263–296. doi: 10.1007/s11211-008-0071-2
- Uhlmann EL, Dasgupta N, Elgueta A, Greenwald AG, Swanson J. (2002). Subgroup prejudice based on skin color among Hispanics in the United States and Latin America. *Social Cognition*, 20, 198–225. doi: 10.1521/soco.20.3.198.21104
- Urban LM, Miller N. (1998). A theoretical analysis of cross-categorization effects: A meta-analysis. *Journal of Personality and Social Psychology*, 74, 894–908. doi: 10.1037/0022-3514.74.4.894
- Van Iddekinge CH, Raymark PH, Roth PL. (2005). Assessing personality with a structured employment interview: Construct-related validity and susceptibility to response inflation. *Journal of Applied Psychology*, 90, 536–552. doi: 10.1037/0021-9010.90.3.536
- Wittenbrink B, Judd CM, Park B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81, 815–827. doi: 10.1037/0022-3514.81.5.815
- Ziegert JC, Hanges PJ. (2005). Employment discrimination: The role of implicit attitudes, motivation and a climate for racial bias. *Journal of Applied Psychology*, 90, 553–562. doi: 10.1037/0021-9010.90.3.553

APPENDIX

Names: Henk, Daan, Jaap (*Dutch first names of men*), Mohammed, Mahmud, Ali (*Arab first names of men*), Janneke, Sanne, Anne (*Dutch first names of women*), Semra, Fatima, Samira (*Arab first names of women*). The following last names were used: De Jong, Janssen, de Vries, van Dijk (*Dutch last names*) and Abdallah, Shadid, Hamoudi, Ozturk, Benmoussa, Hassan, Khadousi (*Arab last names*).

Note that we used the same names for the different job types (as this was a between subjects' factor) and that some names, like Mohammed, were used in both Study 1 and Study 2 (but never in the same combination of first-last name).

Affiliations: Rotterdamse schaakclub (chess club), Kralingse roeibond (rowing club), Rotterdamse tennisvereniging (tennis club), ouderraad van de lokale basisschool (govern body of local school), personeelsvereniging (staff association), Ridderkerks' filmfestival (film organization), milieuvereniging (environmental advocacy group), de ondernemersvereniging (entrepreneurs' association) (*Neutral affiliations*).

Arabische jongerencentrum (Arab youth center), belangenorganisatie voor de Turkse en Marokkaanse werknemers (Arab staff organization), de Turkse oudervereniging (Turkish parents' association), Marokkaanse oudervereniging (Moroccan parents' association), de Arabische belangenvereniging (Arab association), Arab alumni organisatie (Arab alumni association), film vereniging (Arab film club), cultuurgroep (Arabic cultural association) (*Arabic affiliations*).