More on Why Lakisha and Jamal Didn't Get Interviews: Extending Previous Findings Through a Reproducibility Study

William G. Obenauer

William G. Obenauer
Corresponding Author
University of Maine
Maine Business School
168 College Ave.
Orono, ME 04469
william.obenauer@maine.edu

The proper citation for this article is:

Obenauer, W. G. (2023). More on why Lakisha and Jamal didn't get interviews: Extending previous findings through a reproducibility study. Journal of Management Scientific Reports, 1(2), 114–145. https://doi.org/10.1177/27550311231167366

The published journal article can be accessed at:

https://doi.org/10.1177/27550311231167366

MORE ON WHY LAKISHA AND JAMAL DIDN'T GET INTERVIEWS: EXTENDING PREVIOUS FINDINGS THROUGH A REPRODUCIBILITY STUDY

ABSTRACT

In 2004, Bertrand and Mullainathan published an innovative piece of research that involved sending nearly 5,000 fictitious resumes to employers. Their paper is commonly cited for finding that applicants with White-sounding names benefitted more from experience on their resumes and received 50 percent more invitations to interview than other applicants. The current research, however, demonstrates that while Bertrand and Mullainathan made a critically important contribution to the literature on employment discrimination, there is still much more that we can learn from their data. Through a reanalysis of Bertrand and Mullainathan's data, we find that discriminatory effects were stronger in conditions in which a job posting's experience requirements were ambiguous, applicants with first names of Arabic origin experienced higher levels of discrimination than applicants with other non-White-sounding names, and that the discriminatory effects of having an African-American-sounding name could not be empirically differentiated from the discriminatory effects associated with a name's frequency within the overall population. Our findings contribute to the literature on human information processing and offer important practical contributions regarding how employers can potentially reduce discrimination in selection processes. Additionally, we offer important suggestions for the development of racial manipulations in future experimental research.

Keywords: Diversity/gender, information processing, open science, replication studies, research design, selection/staffing, independent reproducibility

INTRODUCTION

The words "racial discrimination is alive and well" were retweeted over 2,500 times after Samia Jalal submitted two separate resumes for an open marketing executive position at a Dublin radio station and received one rejection and one invitation to interview. The resumes were identical with the exception of one thing: the racial association of the applicant's name (Jalal, 2016). Although the station attempted to rationalize the discrepancy as simply an administrative error (Duffy, 2016), this outcome is actually consistent with one Fortune 500 job recruiter's public statement that she would call an applicant named Jennifer before calling an applicant named Shaniqua (ABC News, 2004). This type of workplace discrimination is often considered by people who identify as racial minorities when naming their children (Drayton, 2013) or deciding how to list their names on a resume (Asare, 2020).

The presence of racial discrimination in resume screening has been established in the literature for decades. For example, Bertrand and Mullainathan (2004a; hereafter referred to as B&M) used an innovative correspondence study to find evidence that the demographic traits associated with an applicant's first name influenced the likelihood of an applicant being invited to interview for the position. Specifically, they found that applicants with White-sounding names received 50 percent more invitations to interview than applicants with what they called African-American-sounding names. B&M achieved this using a very rich data set that incorporated 65 different variables, some of which could be decomposed to derive even more information such as additional ethnic associations of names and overall name frequency within the population. In the current research, we seek to identify what more we can learn about discrimination in resume screening from correspondence studies by re-analyzing data from B&M.

Employment discrimination research

Recent research suggests that racial discrimination in the workplace may have decreased the gross domestic product of the United States economy by as much as $16 trillion over the past two decades (Akala, 2020). Extant research has identified employment discrimination in aspects such as selection (Pager & Quillian, 2005), retention (Obenauer & Langer, 2019), performance evaluation (Park & Westphal, 2013), and being targeted with racial slurs (Rosette, Carton, Bowes-Sperry, & Hewlin, 2013). Highly publicized news stories (e.g., Chaffin, 2022; Costello, 2022; Hilliard, 2022) and a list of discrimination cases curated by the U.S. Equal Employment Opportunity Commission (2023) have validated the continued prevalence of this problem.

Experimental research has been used to identify causal factors that contribute to workplace discrimination such as qualification ambiguity (Dovidio & Gaertner, 2000) and prototypicality (Rosette, Leonardelli, & Phillips, 2008). Recently, however, accurately identifying discriminatory effects in experimental research has become more complex. For example, recent replications of Rosette et al. (2008) conducted by two different research teams both failed to find evidence of White leaders receiving more favorable evaluations than their non-White counterparts (Obenauer & Kalsher, 2022; Ubaka, Lu, & Gutierrez, 2022). The misalignment between these recent findings and ongoing evidence of workplace discrimination in the real world suggests that current participants in experimental research may be modifying their behavior in response to perceptions of social desirability.

Kellar and Hall (2022) offered audit studies as a potential solution to the problem of social desirability because respondents in audit studies are unaware that they are participating in research. In traditional audit studies (e.g., King & Ahmad, 2010; Neumark et al., 1996; Pager & Quillian, 2005), researchers have multiple actors apply for jobs and then analyze relationships between candidates' demographic traits and follow up calls received from employers (hereafter referred to as callbacks). B&M argued that the correspondence study, a type of audit study where researchers send fictitious resumes or applications in response to job openings, is the ideal type of audit study as it reduces the likelihood that an experimenter's bias will influence findings.

Audit studies, however, come at considerable costs to researchers and society. The researcher incurs costs associated with running the experiment and data management. The cost to society of any one audit study is negligible, but one must consider the potential cost of audit studies should research trend towards that methodology. If each one of the 146 American institutions listed as engaging in "very high research activity" by The American Council on Education (ACENET, 2022) was responsible for leading only five new correspondence studies annually and the average correspondence study submitted 5,000 resumes in response to posted job openings, these studies would account for more than 3.6 million fictitious resumes entering the employment marketplace. To minimize this impact on the workplace, researchers have a responsibility to extract as much insight as possible from the rich data sets associated with each correspondence study. That is the purpose of the current research.

The current research

The current research constitutes a series of direct and constructive independent reproducibility studies that also examine the generalizability of B&M's findings through the testing of new relationships. B&M serves as an appropriate target for reproduction because of its high impact and relevance to current societal issues. In addition to having more than 6,600 citations on Google Scholar as of February 2023, this paper has influenced the development of academic theory regarding discrimination in selection decisions (e.g. Charles & Guryan 2008; Umphress, Simmons, Boswell, & Triana 2008), and is frequently referenced by the mainstream media (e.g., Fussell 2016; Parkinson & Smith-Walters 2015). The continued prominence of this study's findings has even contributed toward public policy through movements to eliminate visibility of candidate names from employment applications (e.g., Prime Minister's Office, 2015).

The current research begins with a direct reproduction of B&M through a reanalysis of their data using the comparable methods. We then perform a series of constructive reproductions where we examine the impact of a name's origins, name frequency within the population, and an applicant's gender on callback likelihood. Our results show that B&M's primary findings are very robust to model specification. We offer the additional insights, however, that resumes with names of Arabic origin had lower callback rates than those with other non-White-sounding names, the penalty (in terms of callback rate) for having an African-American-sounding name could not be distinguished from the penalty for having a name that is uncommon within the overall population, and that these discriminatory effects were stronger when posted job qualifications were ambiguous.

THEORETICAL BACKGROUND AND HYPOTHESES

Background of Bertrand and Mullainathan's (2004) Study

B&M conducted a correspondence study where they manipulated the race of applicants submitting resumes to job openings. To manipulate the race of a job applicant, B&M used a collection of names that they identified as distinctly White-sounding or African-American-sounding. First names used were identified by collecting name frequency data for all births recorded by the Massachusetts Department of Health (MDOH) from 1974 to 1979. At that time the MDOH birth data only coded race as Black,[1] Chinese, Hispanic, White, other non-White, and unknown, with the latter two categories being used sparingly. B&M calculated the uniqueness of

[1] While B&M used the term African-American in their research, the actual category name listed in the MDOH data at the time was Black.

a name to a particular race by dividing the total number of times that a name was associated with births for that race and a specific gender and dividing that number by the total number of times that the name was associated with all births for that gender. For example, if a name was used 995 times for White males and 1000 times for all males, it was identified as being 99.5 percent unique to White males. This process was completed for White males, White females, African-American males, and African-American females.

The racial implications of names with the highest uniqueness score for each category were then tested through a forced-choice survey on 30 respondents in the Chicago area. The survey asked participants to classify names as "White, African-American, Other, [or] Cannot Tell" (p 995). Because priming participants with the presentation of a choice can restrict their cognitive ability to access alternative options (Herr, 1986; Srull & Wyer, 1979), participants in this validation exercise were less likely to consider alternative racial and ethnic categories such as Asian, Arabic, Jewish, Hispanic, etc. This is relevant because more than 25 percent of the names that B&M used for their African-American condition were names of Arabic origin (Aisha, Hakim, Kareem, Jamal, and Rasheed). Names that survey participants did not readily associate with the intended race (either White or African-American) were eliminated from the experimental manipulation. Although applicant first names were the primary racial manipulation, last names were manipulated such that a series of White-sounding and African-American-sounding last names were also identified by the authors, though the process for this identification was not explicitly outlined in the manuscript.

From July 2001 to May 2002, B&M's research team sent out 4,870 resumes in response to job advertisements in Chicago and Boston. Two to four resumes were sent in response to each advertisement. For each resume with a White-sounding name that was sent to an employer, a resume of equal quality with an African-American-sounding name was also sent to that employer. Resumes included fictitious addresses and real phone numbers that directed callers to voicemail. Some resumes also included email addresses. The researchers collected data on the responses to resume submissions that were received via email or telephone.

B&M reported that White applicants received significantly more callbacks than African-American applicants after submitting a resume for a published opening. They reported that the differential callback rate attributable to race was similar for male and female applicants. They also provided some evidence that White candidates benefitted more than other candidates from employment skills and experience. Finally, B&M reported no evidence of gender discrimination. Although they did find that females had higher callback rates than males for sales positions, this difference was not statistically significant. The current research seeks to extend these findings by examining whether the impact of resume quality on callback rates differs by applicant gender, how the frequency of a name's use within in the general population impacts callback rates for applicants with that name, and whether applicants with first names of Arabic origin experience lower callback rates than other non-White applicants.

Anti-Arab Bias

Individuals of Arabic descent are frequently targets of discrimination that is unique to that experienced by other racial and ethnic minorities. Those who strongly identify with their ethnicity report higher levels of discrimination than those who do not (Awad, 2010). These perceptions of discrimination are validated by both negative portrayals in the American entertainment industry (Shaheen, 2003) and a body of empirical evidence. For example, in a sample of over 500 college students, Bushman and Bonacci (2004) found that individuals of Arabic descent were subject to higher levels of prejudice than members of four other racial/ethnic groups (i.e., White, Black, Hispanic, and Asian) and that this prejudice can lead to adverse actions. Additionally, experimenters responding to roommate request advertisements found that inquiries from people with European-American names received significantly higher response rates than inquiries from people with Arabic names in three out of four American cities (Gaddis & Ghoshal, 2015).

The impact of this anti-Arab prejudice has been felt in the job market as well. Using a study design similar to that of B&M, Widner and Chicoine (2011) found that resumes with Arabic names receive fewer callbacks than resumes with White-sounding names. Results from a resume sorting activity where both name and organizational affiliations were manipulated revealed evidence of a moderate level of discrimination against applicants with Arabic names (Derous, Nguyen, & Ryan, 2009). Through a series of studies that included sending applications to job openings and in-basket sorting exercises, Derous, Ryan, and Nguyen (2012) found evidence of anti-Arab hiring discrimination. Subtle discrimination against applicants whose appearance is consistent with the stereotype of an Arabic background (Suleiman, 1999) has also been found in interpersonal application processes (King & Ahmad, 2010).

Because some of B&M's data collection took place before the tragic events of September 11, 2001 (9-11) unjustifiably impacted expressions of prejudice against Arab-Americans (Semaan, 2014), it is relevant to validate the existence of an anti-Arab prejudice in America prior to 2001. Stockton (1994) outlined the relationship between negative Arabic stereotypes and prejudice against Arab-Americans. Boycotts and oil embargos contributed to the perpetuation of these prejudices by fueling tensions between Americans and Arab nations decades before 9-11 (Kaikati, 1978). Biases and prejudice against those of Arabic descent publicly manifested themselves during the conflicts in the Middle East of the 1990s (Anderson, 1991). At times, prejudice against Arab-Americans became so salient that it even resulted in violence (Lerner, 1986).

Despite the last name component of B&M's racial manipulation incorporating only last names that are *not* of Arabic origin (e.g., Jackson, Washington, Williams), the aforementioned anti-Arab bias is highly relevant to their research. Human information processing theory states that people categorize information about actors and form mental models, or prototypes, in response to previous experiences and exposures (Lord & Maher, 1993; Rosch, 1978). Prototype development and recall takes place through automatic processing, where the person is not engaged in critical thinking or analysis (Lord & Smith, 1983). In organizational contexts, when presented with a stimulus, individuals may subconsciously draw upon categorized prototypes that will influences tasks such as evaluation and selection (Obenauer & Kalsher, 2022; Rosette et al., 2008; Ubaka et al., 2022). Consequently, a first name alone may be sufficient for prototype retrieval. Building on human information processing theory, when presented with a name such as Kareem Robinson, rather than critically debating whether the applicant is of Arab descent (based on first name) or of another race/ethnicity (based on last name), an evaluator engaged in automatic processing is likely to draw upon their prototypes for Arab-Americans and activate relevant biases in response to the signal communicated by the applicant's first name.

Even if the evaluator considers the signals communicated by both the first name and the last name, the potentially conflicting racial and ethnic signals sent by a name like Kareem (Arabic) Robinson (Black/African-American) are unlikely to be problematic as, according to intersectionality theory, people can hold multiple identities at once (Ozturk & Berber, 2022). Although intersectionality theory historically emphasized the coexistence of gender and racial identities (e.g. Arai, Bursell, & Nekby, 2016; Daniels & Thornton, 2019; Salerno, Peter-Hagene,

& Jay, 2019), intersectionality is relevant when people are associated with multiple racial and ethnic groups. Furthermore, the double jeopardy component to intersectionality theory states that discrimination may be cumulative when individuals are punished for holding multiple stigmatized identities (Cortina, Kabat-Farr, Leskinen, Huerta, & Magley, 2013). Therefore, in the case of a person who is given an Arabic first name and a last name that sounds Black/African-American, that individual may experience cumulative discriminatory effects for each component of their racial and/or ethnic identity. Integrating this literature with the structure of B&M's data set, we propose to test the following hypothesis.

Hypothesis 1: Applicants will experience differential callback rates attributable to first names of Arabic origin that are distinct from differential callback rates attributable to non-White names in general such that applicants with first names of Arabic origin will receive lower callback rates than all other applicants.

Familiarity bias

First names play a much greater role in the development of perceptions than simply conveying the probable race of an individual. They have been shown to influence perceptions of intelligence (Young, Kennedy, Newhouse, Browne, & Thiessen, 1993), impressions of creativity (Lebuda & Karwowski, 2013), physical attraction (Erwin, 1993), and perceived success (Mehrabian & Piercy, 2001). Factors such as the spelling or length of a name can play an important role in how people form impressions (Mehrabian, 2001). Also, the use of a nickname instead of a birth name (e.g., Katie vs. Katherine) can influence perceptions (Mehrabian & Piercy, 2001), providing reason to believe that evaluators may respond differently to resumes with different names that are each associated with the same race or ethnicity.

One factor related to first names that is highly relevant to B&M's work and has been shown to play an important role in impression formation is familiarity bias. In a sample of college students, Young et al. (1993) found that individuals with less common names were perceived as less popular and of lower intelligence. Cotton, O'Neill, and Griffin (2008) found that not only are common names more well-liked than unique names, but that commonality of a name played a larger role in influencing hiring decisions than the ethnic or racial association of the name did.

B&M recognized potential importance of name commonality, but their discussion on this issue was limited to addressing the commonality of a name within a group of other names that are associated with the same race and gender. We argue, however, that this type of relative frequency is of limited value when trying to understand the potential impact of familiarity bias because to be familiar is to be "easy to recognize because previously experienced," (Cambridge University Press 2016). This previous experience is not likely to be influenced by how a name is categorized. For example, from 1974 to 1979, the name Kristen was given to female babies in Massachusetts 37 times as frequently as the name Latoya. Latoya was the most popular name that B&M classified as an African-American name within their name bank, whereas Kristen ranked fourth amongst White-sounding female names. The fact that Latoya had a higher relative frequency than Kristen does not make general members of the population more familiar with the name Latoya. Instead, the higher frequency of use of the name Kristen within the overall population is likely to influence people's familiarity with the name.

The overall frequency of a name's use within the population (as opposed to relative frequency) is also likely to influence associations that occur within the classification processes associated with automatic processing (Lord & Maher, 1993; Lord & Smith, 1983). For example, in terms of probability, it is more likely that successful past hires will be associated with more frequently occurring names than less frequently occurring names. Consequently, applicants with more frequently occurring names may appear to be more prototypical of the commonly successful employee than applicants with less frequently occurring names (Rosch, 1978). To the extent that talent evaluators are engaged in automatic processing while attempting to satisfy multiple demands, this prototypicality can provide an advantage to applicants with frequently occurring names. To address this, we propose to test the following hypothesis.

Hypothesis 2: Applicants will experience differential callback rates attributable to name frequency within the population that are distinct from differential callback rates attributable to a name's racial and/or ethnic associations such that applicants with less frequently used names will receive fewer callbacks than applicants with more frequently used names.

Gender Gap

In addition to learning about the impact of applicant race and ethnicity on employment selection, it's possible that the richness of B&M's data allows us to learn more about gender discrimination as well. Although B&M did not find evidence that applicant gender directly influenced callback rates, and they found minimal evidence that racial differences in callback rates varied by applicant gender, these insights do not preclude the possibility of gender effects being present within their research. A large body of organizational research has demonstrated that females face unique challenges in the workplace that are often contextual in nature.

For example despite equal pay laws, females frequently earn less than men in similar positions (Castilla, 2015; Judge & Livingston, 2008; Newton & Simutin, 2015). A phenomenon known as the glass ceiling describes the barriers females often face when climbing the organizational ladder (Cotter, Hermsen, Ovadia, & Vanneman, 2001; Goodman, Fields, & Blum, 2003; Wright & Baxter, 2000). Females are also more likely to be promoted into precarious leadership positions (Haslam & Ryan, 2008; Ryan & Haslam, 2005; Ryan, Haslam, & Postmes, 2007). Evidence of gender discrimination is not limited to leadership, however, as research has used a variety of methods to quantify gender discrimination in hiring (e.g., Kübler et al., 2018; Neumark et al., 1996; Pratto et al., 1997; Uhlmann & Cohen, 2007).

The above description is not an exhaustive list of the types of workplace discrimination faced by females, but it illustrates why B&M's lack of evidence for gender discrimination within their research may be somewhat surprising. These findings, however, are not completely unique as other correspondence studies have found conditions in which female candidates receive a higher number of callbacks than their male counterparts (Arai et al., 2016; Booth & Leigh, 2010). To contextualize B&M's findings within the tension in the literature, we consider a component to their experimental design that complicated identifying gender effects. Males and females of similar racial characteristics and resume quality were never directly compared by B&M. Their data structure, however, allows for analysis of resume quality effects by gender.

The shifting standards literature has shown that evaluators sometimes set lower standards for members of groups that have been historically treated as lower status, such as women (Biernat & Kobrynowicz, 1997). This is because on the lower end of the evaluation scale, members of traditionally marginalized groups tend to be compared to other members of their group, within the context of negative stereotypes. For example, in an environment where females are frequently targets of discrimination, a female applicant may be perceived as well-qualified for a female, rather than being evaluated within the context of the overall pool of applicants (Biernat & Thompson, 2002). If such a context resulted in a lower minimum threshold for a female applicant to receive an invitation to interview, a male applicant with a low quality resume could be less likely to receive an interview invitation than a female applicant with a similarly low quality resume (Biernat & Kobrynowicz, 1997). This statement is consistent with Biernat and Fuegen's (2001) argument that despite having a lower probability of receiving a job offer, female job applicants are more likely to be short-listed for a job than their male counterparts.

This shift in standards, however, may not benefit females when qualifications are high. The gender discrimination described above limits opportunities for females to hold visible positions in the workplace. Using human information processing theory to build on Rosette et al.'s (2008) argument that demographic traits are incorporated into prototypes for successful employees, one might expect that a male's high-quality resume would signal consistency with the prototype for a successful male employee, therefore resulting in a more favorable outcome for a male applicant with a high-quality resume. Such an assumption would be consistent with recent findings that high-performing females were evaluated as having lower potential than their male counterparts (Benson, Li, & Shue, 2022). Collectively, if lower quality female resumes are more likely to be selected than lower quality male resumes and higher quality male resumes are more likely to be selected than high quality female resumes, it would appear that male applicants may benefit more from resume quality than female applicants.

Human information processing theory, however, could also provide us with a pathway to theorizing a conflicting effect. An alternative expectation may be that because of its inconsistency with negative gender stereotypes, a high-quality resume submitted by a female applicant may be notable such that it stimulates evaluators to break the automatic processing script and engage in more controlled processing during the evaluation process, thus reducing the impact of gender bias in selection. Several studies support this argument. For example, Biernat and Vescio (2002) found that a selection preference for males was only present when qualifications were ambiguous, thus leaving evaluators more likely to engage in automatic processing. Boldry, Wood, and Kashy (2001) found that male military cadets were evaluated more favorably than female cadets on key metrics such as motivation and leadership, but that the gap was smaller for highly qualified cadets than it was for average cadets. Similarly, Koch, D'Mello, and Sackett's (2015) meta-analysis found decreased gender bias when females were highly qualified or evaluators were motivated to engage in careful, controlled analysis.

Consequently, it appears that strong qualifications may benefit female applicants disproportionately. Furthermore, Obenauer and Kalsher (2022) suggested that when engaged in controlled processing, it's possible that corrections for bias *may* be imperfect such that traditionally marginalized candidates *could be* evaluated more favorably. In such a case, this *could* result in females benefiting more from resume quality than their male counterparts. Consequently, we propose testing the following competing hypotheses:

Hypothesis 3A: Applicants will experience differential callback rates attributable to resume quality such that male applicants will experience a greater benefit from resume quality than female applicants experience.

Hypothesis 3B: Applicants will experience differential callback rates attributable to resume quality such that female applicants will experience a greater benefit from resume quality than male applicants experience.

METHODS

Data transparency

New data constructed for these analyses, a data dictionary, Stata syntax, a full description of the direct reproduction, and additional tables and analyses are available in an associated Open Science Framework (OSF) project.[2]

Data

Primary analyses. The primary data for analyses were retrieved directly from the target publication's supplementary materials (Bertrand & Mullainathan, 2004b). These data included 4,870 observations for submitted resumes that varied based on applicant race (White-sounding name=2,435; not-White-sounding name=2,435), applicant gender (male=1,124; female=3,746), resume quality (high quality=2,446; low quality=2,424), and type of position applied for (administrative=2,792; sales=2,078). Three-hundred ninety-two submitted resumes (8.049 percent) resulted in callbacks. Data on omitted variables were not available (M. Bertrand, personal communication, March 8, 2022).

Frequency analyses. B&M's published dataset did not include the name frequency that was derived from Massachusetts birth records. To obtain these data, birth records for 1974 through 1979 were obtained directly from the Massachusetts Department of Public Health.

Measures

Listed below are the dependent variable for this study, key applicant demographic variables, and new variables introduced for this reproduction.

Callback. This binary variable (1=applicant received callback, 0=applicant did not receive callback) serves as the dependent variable for analyses.

Not-White applicant name. This variable is derived from B&M's *race* variable such that an applicant name that was identified as a common African-American name in the target study

[2] An OSF project that includes data, a data dictionary, analyses syntax, and additional reporting of results can be accessed at https://osf.io/jq5f9/?view_only=921284f626d849979ca5b6871c0a9bd6

was classified as "not-White" in the current research (1=applicant name was identified as a common African-American name by B&M, 0=name was identified as a common White name by B&M).

Female applicant name. This variable is derived from B&M's *sex* variable (1=name more commonly associated with females, 0= name more commonly associated with males).

Arabic applicant name. This is a binary variable indicating whether a name was of Arabic origin (1=name is of Arabic origin, 0=name is not of Arabic origin). Five names from the original data (Aisha, Hakim, Kareem, Jamal, and Rasheed) have been identified by linguists as being of Arabic origin (Hanks, Hardcastle, & Hodges, 2006) and were coded as such.

To validate that distinctly classifying names of Arabic origin from other non-White names was appropriate, we examined MDOH data from 2011 to 2016. Beginning in 2011, MDOH began using 50 different racial and ethnic classifications and allowed a birth to be classified as being associated with multiple categories of race and ethnicity (Massachusetts Department of Public Health, personal communication, October 24, 2018). Birth records for 2011 through 2016 were obtained directly from MDOH to examine the races and ethnicities associated with names of Arabic origin.

We calculated the percentage of births associated with a specific name and the racial classification of Black. For privacy purposes, the MDOH data codes the number of occurrences of a classification as -999 if the number of occurrences in a year is greater than zero and less than five.[3] This means that we did not have complete data on the number of occurrences for each name. Consequently, we calculated mean ranges. At the high end of the range, screened data were treated as a value of four. At the low end of the range, screened data were treated as a value

[3] Due to the low number of occurrences of the non-White names used by B&M in the MDOH data, the MDOH restriction on data reporting prevented us from examining several other relevant racial and ethnic categories.

of one. We then used these data to calculate weighted averages for the names B&M used for their non-White applicant resumes based upon whether or not the names were of Arabic origin. These averages showed in the 2011-2016 MDOH data, 50.000 to 78.571 percent of the births in which a child was given a non-Arabic first name that was used by B&M for their non-White applicant condition were associated with at least one parent who identified as Black. When a child was given an Arabic first name that was used by B&M for their non-White applicant condition, this range was reduced such that only 30.682 to 51.136 percent of the births were associated with at least one parent who identified as Black. This difference in ranges suggests that in addition to having a different historical origination, names of Arabic origin have different demographic associations than the other names that B&M used to represent not-White job applicants. This insight further supports the decision to categorize names of Arabic origin distinctly from those of non-Arabic origin.

Name frequency. This variable represents the log normal of the total number of times that a first name appeared in the MDOH data from 1974 to 1979. Details on variable construction are included in the online data dictionary.

Data analyses

Direct reproduction. Because the analyses described in this section were conducted with the original data, we refer to these analyses as direct reproduction (Köhler & Cortina, 2021). Successfully conducting direct reproductions was a necessary first step to this research as it demonstrates that variability observed in subsequent constructive reproductions should not be attributed to unintentional variations in analytical procedures. All direct reproductions were conducted using the same statistical techniques described in the text by B&M. B&M did not provide the analytical code used to perform statistical analysis so it is possible that there could be minor differences in statistical techniques related to code specification or software updates. However, our consistent ability to reproduce their results indicates that our methodological analyses were similar to those used in the target study.

Test of proportion. Tests of proportion were conducted using Stata17's *prtest* command. Test of proportion were used for direct reproductions, constructive reproductions, and testing of Hypotheses 1 and 3.

Probit regression. Probit regressions were conducted using Stata17's *probit* command with the *vce(cl [clustervariable])* option, with standard errors clustered on the variable *adid.* Marginal effects for each variable were calculated using Stata17's *margins* command with the *dydx(*)* option after storing regression estimates. Probit regressions were used for direct reproductions, constructive reproductions, and testing of all hypotheses.

Linear probability model. The linear probability model (LPM) is an appropriate robustness test to address concerns about the appropriateness of using interactions within the restrictions imposed by logistic regressions (Ai & Norton, 2003). LPM results are often similar to those of binary logistic regression, providing clarity as to how interaction results can be interpreted (Chatla & Shmueli, 2016; Obenauer & Langer, 2019). LPM regressions were conducted using the *reg* command in Stata17 with the *vce(cl [clustervariable])*. LPM was used exclusively for robustness tests and all reported coefficients are unstandardized.

RESULTS

Direct reproduction

The results of direct reproductions generally mirrored those of B&M. Below, we discuss only where our findings meaningfully diverge from those of B&M. Discrepancies that are likely attributable to rounding or transcription errors, are not discussed here. A full description of the direct reproduction is available in the associated OSF project (see Footnote 2 for link).

Insert Table 1 About Here

Our findings diverged from those of B&M in two places when attempting to reproduce the "All resumes" column of their Table 5. This column showed the impact of various resume characteristics on the likelihood of receiving a callback for all applicants. B&M reported that the presence of an email address on a resume on callback likelihood was significant. They also reported that computer skills impacted callback likelihood, although they did not specify if this impact was significant. The p-values for email ($p= .152$) and computer skills ($p= .089$) in the current research, however, did not meet the threshold for statistical significance.

B&M's Table 6 (see the Direct Reproduction file in the associated OSF project) reported the results of probit regressions analyzing the impact of an applicant's neighborhood characteristics on the likelihood of receiving a callback. B&M reported that "applicants living in Whiter, more educated, or higher-income neighborhoods have a higher probability of receiving a callback" (p. 1003). Consistent with this interpretation we found that the marginal effects of a neighborhood's education level and per capita income on callback likelihood were significant ($p= .016$ *and* $p= .009$, *respectively*). We also successfully reproduced the marginal effects of a neighborhood's racial composition on callback likelihood and the associated standard errors reported by B&M (and by extension, the test statistics), but our results indicate that the referenced marginal effects were not significant ($p= 0.101$). For the level of significance to approach the cutoff value of $p<0.05$, the test statistic would have to be interpreted using a one-tail test. This would not be consistent with how the results of a probit regression are typically interpreted and reported, and the 95 percent confidence interval for the marginal effect would still include zero. Consequently, the current analyses do not support B&M's claim that the racial composition of an applicant's neighborhood was related to the applicant's likelihood of receiving a callback.

Relevant Exploratory Analyses

Interaction of race and resume quality. To better understand how resume quality and applicant race may interact to influence callbacks, we performed tests of proportion on callback by race within subsamples of high- and low-quality resumes (see *Table 1 Models J and K*). Resumes with White-sounding names had a higher callback rate in both the high-quality resume ($W=10.793\%$, $NW=6.705\%$, $z=3.580$, $p<0.001$) and the low-quality resume ($W=8.498\%$, $NW=6.188\%$, $z=2.180$, $p= .029$) conditions. These results indicate that applicants with White-sounding names experienced an advantage irrespective of resume quality.

Because interaction terms can be used to provide a more informative understanding of moderation effects (Stone-Romero & Anderson, 1994), we then regressed the callback dummy on applicant race, resume quality, and the interaction of race and resume quality using a probit model with standard errors adjusted for clustering at the advertisement level. Model A of *Table 2* shows that the marginal effect of the interaction of race and resume quality was not significant ($p= .204$). Results using LPM were similar. Collectively, these findings indicate that the impact of resume quality on callback likelihood was not strongly influenced by applicant race.

Differential effects of resume characteristics. B&M's Table 5 showed the results of probit regressions analyzing the impact of an applicant's race and a variety of resume characteristics on the likelihood of receiving a callback. B&M reported that applicants with White-sounding names were more strongly impacted by characteristics of their resumes. This interpretation was based on subsample analyses. Subsample analyses, however, do not test the significance of differences between samples.

Figure 1 shows that the 95-percent confidence intervals for the marginal effects of resume characteristics by applicant race overlap for every characteristic, indicating that these differences may not be significant. For robustness, Model F of *Table 2* reports the results of a probit regression where the binary variable for race was interacted with different resume characteristics. The main effect of race was not significant ($p= .485$), but the model indicates that candidates with White-sounding names benefitted more from experience than those with non-White-sounding names. Results from LPM were similar. These results also indicate that while applicants with White-sounding names appear to have benefitted more from experience than other job applicants, no other resume characteristics had significant differential effects on the likelihood of an applicant receiving a callback.

Job requirements (robustness test). An LPM model examining the impact of any published job requirement on the callback likelihood showed a significant positive coefficient on the interaction of applicant race and requirement ($\beta=0.033$, $SE=0.016$, $p= .043$). Because the coefficient for the main effect of the presence of a job requirement was negative and significant ($\beta=-0.044$, $SE=0.019$, $p= .023$), the interaction term indicates that the negative effect of a published job requirement on the likelihood of a callback was *weaker* for non-White applicants than it was for White applicants. This result differed from that of the probit regression.

To reconcile the difference in results pertaining to the impact of the interaction of job requirements and applicant race in the probit and LPM regressions, we conducted a series of tests of proportions (*see Table 1, Models L through O*). Models L and M show that whether or not job requirements were present in the advertisement, the White candidate received a higher proportion of callbacks than the non-White candidate (*p= .002 and p= .003, respectively*). Models N and O, show, however, while the White candidate received a higher portion of callbacks when no job requirements were posted (*13.127%*) than when job requirements were posted (*8.712%, p= .003*), callbacks received by the non-White applicant were not influenced by the presence of job requirements (*p= .354*). This series of proportion tests is consistent with the results of the LPM showing that the interaction effect of applicant race and job requirements on callbacks was significant.

Posted requirements and applicant qualifications. Finally, to better understand the limited impact of applicant qualifications and job requirements on an applicant's likelihood of receiving a callback, we considered that neither qualifications nor requirements occur in a vacuum. Theoretically, qualifications should matter such that they serve to meet a job requirement. A comparison of applicants' qualifications and the requirements of the jobs they applied for showed that 98.446 percent of resumes submitted met the posted experience requirements, 91.538 percent met education requirements, 94.786 percent met posted computer skills requirements, and 94.580 percent met all posted requirements that could be verified against the resume based upon available data.

Given that applicant experience was the one variable that differentially impacted an applicant's likelihood of receiving a callback, our exploratory analyses focus on this variable. The fact that over 98 percent of applicants met experience requirements when they were included in a job advertisement restricted us from testing the impact of meeting requirements versus failing to meet them, but in a practical sense, it provided us with two relatively balanced conditions to examine: employment competitions in which applicants' experience was subjectively evaluated (no experience requirements posted) and employment competitions in which applicants' experience could be objectively compared to posted job requirements (experience requirements posted). To examine how ambiguity in experience requirements would influence the differential effects of applicant experience that were previously reported, a subsample analysis of *Table 2's Model F* was conducted.

The subsample analysis showed that when a job posting did not include experience requirements (*see Table 2, Model G*), the marginal effect of years of experiences was positive and significant (*marginal effect=0.017, SE=0.005, p<0.001*). Consistent with Model F, the interaction of race and experience had a negative and significant marginal effect (*marginal effect= -0.014, SE=0.007, p= .027*), indicating that non-White applicants experienced a considerably smaller benefit from experience than their White counterparts. Model H shows that when experience requirements were shared in the job posting, neither the marginal effect of experience (*p= .326*) nor the interaction of experience and applicant race (*p= .797*) were significant. Results from LPM regressions were consistent with those of the probit model.

It should be noted, however, that the presence of experience requirements did not fully eliminate racial discrimination from the selection process. When attempting to reproduce the results of Model C in the subsample of observations where job postings included experience requirements, the marginal effect of applicant race on the likelihood of receiving a callback was reduced, but remained significant (*marginal effect= -0.023, SE=0.009, p= .009*).

H1: Anti-Arab prejudice

Hypothesis 1 stated that applicants with names of Arabic origin would experience discrimination effects beyond those experienced by other non-White job applicants. In the first test of this hypothesis, we compared callback rates for non-White applicants with first names of Arabic origin to those of non-White applicants with non-Arabic first names using a test of proportion. As shown in *Table 1, Model P,* non-White applicants with non-Arabic first names received 1.874 times as many callbacks as those with names of Arabic origin. The difference in callback rate of 3.275% was statistically significant (*SE=0.011*, *z=2.502*, *p= .012*).

To contextualize this finding within the overall results of this research, we compared callback rates for applicants with names of Arabic origin to those of applicants with White-sounding names (*see Table 1*, *Model Q*). The results showed that applicants with White-sounding names received 2.576 times as many callbacks as those with Arabic names (*diff=5.904%, SE=0.011*, *z=3.978, p<0.001*). We then reexamined this relationship while restricting the sample to observations in which an applicant with a White-sounding name could be directly compared to an applicant with an Arabic first name responding to the same job advertisement with a resume of similar quality (*see Table 1, Model R*). The results again indicated that applicants with White-sounding names received 2.313 times as many callbacks as applicants with an Arabic first name (*diff=4.918%, SE=0.016, z=2.979, p= .003*). These findings show that the callback gap between applicants with White-sounding names and applicants with Arabic first names was consistently larger than the callback gap between applicants with White-sounding names and other applicants with non-White-sounding names, therefore providing additional evidence that applicants with Arabic first names experienced a greater disadvantage in the job market than other applicants with non-White-sounding names.

Finally, we executed the probit regressions described above, including a binary variable for names of Arabic origin. Model I of *Table 2* shows that consistent with other findings, applicants with non-White names experience a disadvantage in terms of callback rates (*marginal effect=-0.026, SE=0.007, p<0.001*). Applicants with a first name of Arabic origin, however, experience an additional disadvantage (*marginal effect=-0.029, SE=0.013, p= .027*), indicating that the penalty experienced in terms of callback rates by applicants with Arabic names was twice that of other applicants with non-White names. Collectively, these tests show support for Hypothesis 1 which stated that applicants with names of Arabic origin would experience disadvantages that were distinct from those experienced by other applicants with non-White names.

H2: Name familiarity bias

Hypothesis 2 stated that the frequency of a first name within the overall population would influence an applicant's likelihood of receiving a callback beyond any effect attributable to the applicant's race or ethnicity. We began by conducting a rank-order correlation between name-specific callback rate and name frequency within the overall population, using a procedure similar to what B&M described on p. 1009, but focusing on frequency within the overall population rather than gender-race groups. The findings showed a significant correlation between callback rate and name frequency such that more frequently occurring names had higher callback rates ($r= -.534, p<0.001$).

Next, we reproduced the analysis of *Table 2's* Model I using a probit regression that introduced the log of name frequency to the model. The effect of Arabic origin of name on the likelihood of a callback remained significant (*marginal effect=-0.029, SE=0.013, p= .028*), but neither the overall effect of a non-White name ($p= .232$) nor the effect of frequency ($p= .893$) was significant (*see Table 2, Model J*). These effects remained insignificant when removing the variable for Arabic origin from the model. One possible reason for the insignificant marginal effects is that introducing the frequency variable to the model resulted in multicollinearity as a point biserial correlation indicated that the frequency variable and race variable were highly correlated ($r= -.927, t(4868)=170, p<0.001$).

To account for potential multicollinearity, Model K incorporates the name frequency variable into the probit regression without including the primary race variable. In this model, both the effect of Arabic origin of name (*marginal effect=-0.030, SE=0.013, p= .018*) and the population frequency of the name (*marginal effect=0.005, SE=0.002, p= .001*) had a significant relationship with the likelihood of a callback. These findings show partial support for Hypothesis 2 as the observed effect of name frequency on the likelihood of receiving a callback was distinct from that of an Arabic name but was not distinct from our primary race variable.

H3: Gender effects

Hypothesis 3 offered competing hypotheses such that a) male applicants would benefit more from resume quality than female applicants, or b) female applicants would benefit from resume quality more than male applicants. To test these hypotheses, we began by examining the difference in callback rates attributable to resume quality for females. A test of proportions showed that females with high-quality resumes received 1.264 times as many callbacks as those with low-quality resumes (*diff=1.923%, SE=0.009, z=2.139, p= .033*). A similar test showed that the relationship between resume quality and callbacks was not significant for males (*p= .843; see Table 1, Models S and T*). These tests, however, included observations where the gender associated with resumes sent to the same job was not always held constant. For robustness, we re-ran these proportion tests restricting the samples to observations where the gender associated with resumes sent for a specific job posting was held constant for each race group[4]. As shown in Models U and V of *Table 1*, these results were similar to those reported above, providing support for *Hypothesis 3B.*

[4] In other words, all resumes with White-sounding names that were submitted for a specific job posting were given a name associated with the same gender and all resumes with non-White names that were submitted for a specific job posting were given a name associated with the same gender

The analyses described above, however, were imbalanced by position type. For example, 86.463 percent of the female resumes included in Model U were sent to administrative job openings whereas 88.933 percent of the male resumes included in Model V were sent to sales job openings. To address this limitation, we reproduced the analyses used in Models U and V, restricting the sample to resumes sent in response to job openings in sales.[5] As shown in Models W and X of *Table 1,* when restricting analyses to sales positions, the relationship between resume quality and callbacks was neither significant for females (*p= .674*) nor males (*p= .857*). As a final robustness test, we regressed the callback dummy on applicant race, resume quality, applicant gender, and the interaction of gender and resume quality using a probit regression with standard errors adjusted for clustering at the advertisement level. Neither the marginal effect of gender (*p= .858*) nor the interaction of gender and resume quality (*p= .184*) were significant, demonstrating that the previously shown support for *Hypothesis 3B* is limited (*see Table 2, Model L*).

DISCUSSION

Summary of findings

The current research investigated differential outcomes in employment selection and generally successfully reproduced the findings reported by B&M. Specifically, this reproduction research shows that B&M's finding that White applicants received higher callback rates than non-White applicants is robust to a variety of model specifications. We also reproduced their finding that White applicants benefited more from work experience, in terms of callback rates, than non-White applicants. Our primary divergence from B&M is that our analyses did not indicate that having an email address was a reliable predictor of job market performance, but this

[5] There were not enough male resumes sent to administrative openings to conduct subsample analyses based on that type of position.

divergence may be of minimal practical significance as email use is considerably more popular now than it was at the time of the original study.

Our reproductions offer the following new insights from B&M's dataset. First, we found that the differential influence of experience on applicant callback rates was not present in conditions in which the required experience was clearly articulated in the job advertisement. Second, our analyses showed that applicants with Arabic first names received lower callback rates than other non-White applicants. Next, we found that the impact of an applicant's race or ethnicity on the likelihood of receiving a callback could only be differentiated from the impact of a name's frequency in the overall population for applicants with Arabic names. For applicants with non-White-sounding names that were not of Arabic origin, the effect of race could not be distinguished from that of name frequency. Finally, we found that female applicants benefitted more from higher resume quality than male applicants, but this finding was not robust to model specifications in which all applications were submitted for the same type of position.

Theoretical implications

Building on Lord and Maher's (1993) work on information processing, prior research has argued that discriminatory outcomes are most likely to occur when evaluators draw on previously developed prototypes in the evaluation process (Rosette et al., 2008). Recent research has attempted to explain null findings in discrimination research by arguing that when evaluators are engaged in controlled processing, their reliance upon prototypes is reduced (e.g., Obenauer & Kalsher, 2022), but such explanations have been speculative in nature. The current research demonstrates that in a context where evaluators were motivated to engage in controlled processing due to the presence of objective evaluation criteria (i.e., when job requirements were clearly listed), discriminatory effects were weaker than in contexts where no such criteria were published and evaluators may have relied more on automatic processing. This finding makes an important contribution to the application of information processing theory within the realm of employment discrimination research.

Our findings pertaining to discrimination against applicants with Arabic first names also makes an important contribution to this theoretical application. Although the origins of first names identified as Arabic were verified and these names are also frequently associated with the religion of Islam (Hanson & Hawley, 2011), when combined with non-Arabic last names, one may expect the effects of an anti-Arab bias to disappear. According to information processing theory, however, once a prototype has been established, its recall and application takes place through automatic processing, where critical evaluation of additional information is minimized. Our findings provide evidence of this outcome as they suggest that a first name alone may have been enough to trigger the recall and application of a prototype associated with anti-Arab (Bushman & Bonacci, 2004) or anti-Muslin (Saad, 2006) biases.

Finally, our finding that the effects of name frequency could not be differentiated from those of racial associations is a reminder that in any given evaluation, multiple prototypes may be at play. In some cases, prototypes may be similar such that their effects cannot be easily differentiated. In other cases, their effects may be different such that they may appear to counteract each other. Future research on information processing in employment discrimination should continue to explore how signaling similarity with multiple prototypes might influence applicant and/or employee outcomes.

Methodological implications

Reanalysis of data plays an important role in refining methodology for future research (e.g., Blanton et al., 2009; McWilliams & Siegel, 1997; Stauffer & Buckley, 2005). This is of particular importance when the methodology used in a study has a profound impact on subsequent research. B&M has had such an impact as diversity scholars have frequently utilized the names identified by B&M as racial manipulations in their own research (e.g., Friedman et al., 2006; Lee et al., 2015; Milkman et al., 2015; Oreopoulos, 2011b; Rosette et al., 2008; Tadmor et al., 2012; Zapata et al., 2016). Even correspondence studies that have not drawn from B&M's name bank often follow their template for experimental design (e.g., Carlsson & Rooth 2007; Pager & Quillian 2005). These studies frequently validate racial associations with the names utilized, but few have addressed additional information communicated by first names. Our findings suggest that best practices for using first names as racial manipulations in future research would be to 1) choose names with similar *overall* frequencies in the population, and 2) test names for additional relevant demographic associations.

A possible rebuttal to the first recommendation would be to say that because non-Hispanic Whites make up the majority of the population in the United States, White-sounding names will have a much higher frequency in the overall population than non-White-sounding names. This argument, however, is quickly discredited by an analysis of the birth name data from the State of Massachusetts (DPH, 1979).

To illustrate this point, we searched for first names that were used for Black babies 90 percent of the time or more. We took the 9 most common names that met this criterion for each gender and recorded them in Table 3 under the heading "Alternative." As denoted by gray shading, 13 of these 18 names were used in the original study. We then searched for first names that were used for White babies 90 percent of the time or more and whose frequency of use matched that of the previously selected Black names and used them to select an alternative name bank for distinctly White names. This exercise illustrates that although a name could be categorized as both common and distinctly White, there is a difference between distinctly White names and common names. The names shown in Table 3 provide a potential starting point for researchers in following the recommendations outlined above.

Additionally, we observed that some of B&M's "African-American-sounding" names were of Arabic origin. This observation was supplemented by the finding that discriminatory effects experienced by applicants with what they called "African-American-sounding" names differed based upon whether or not the name was of Arabic origin. This finding suggests that when validating names used for racial manipulations in the future, researchers should consider using open-ended questions or more extensive lists of choices to prevent priming effects.

Practical implications

In response to recent calls for research that focuses on actionable steps to achieving equity in organizations (e.g. Nkomo et al., 2019), the current research appears to identify an important actionable step that employers can take in order to make selection processes more equitable. Specifically, our finding that listing experience requirements *reduced* discriminatory outcomes suggests that employers can reduce discrimination in selection processes by setting objective criteria to compare against candidates' resumes or applications. It should be noted, however, that this finding was restricted to objective measures of past experience. Our analyses provide no evidence that criteria based on subjective traits (e.g., "ability to" or "knowledge of") will have similar effects.

This suggestion is also supported by this research's findings that multiple biases may influence selection processes. Consequently, internal attempts to control the effect of biases may be a starting point but are likely an insufficient solution to employment discrimination as hiring agents are unlikely to be aware of *all* of the biases they need to control. Instead, employers should attempt to structure selection processes such that they are less susceptible to bias. For example, employers should consider using name-blind application screening processes (also called anonymous application procedures). Extant research has shown that when application procedures are anonymous, discriminatory effects can be minimized (Åslund & Skans, 2012), likely because name-blind applications limit available stimuli that can activate biases. Anonymous applications also limit the evaluator's ability to engage in cybervetting, which can have additional discriminatory impacts (Wilcox, Damarin, & McDonald, 2022). Name-blind application processes are, however, an imperfect solution (O'Connor, 2016) as application information other than names can signal demographic characteristics and bias the selection process (Foley & Williamson, 2018). Consequently, while such practices may be a good initial first step towards equitable selection, they should not be perceived a definitive solution.

Limitations and directions for future research

One limitation of this research pertains to the impact of the September 11 attacks on our findings related to names of Arabic origin as this event impacted prejudice against Arab-Americans (Semaan, 2014). B&M addressed this issue as this event occurred in the midst of their data collection efforts. They reported that racial differences in callback rates did not change following this event. Their definition of racial differences, however, did not differentiate applicant names based upon whether the name was of Arabic origin. Because additional data on applications (e.g., submission date) were not available we were unable to test this alternative explanation for our finding regarding callback rates for applicants with Arabic names.

Additionally, this study does not account for differences in the individuals who were actually evaluating resumes or the evaluation process. Because of this, we do not know what role characteristics of the evaluator played in these outcomes. For example, extant research has identified implicit associations (Rooth, 2010), dominant groups' desire to maintain their social status (Umphress et al., 2008), concerns about confirming negative stereotypes (Lewis & Sherman, 2003), and perceptions about the contextual needs for the position (Gündemir, Carton, & Homan, 2019) as factors that influence racial discrimination in the hiring process. Future correspondence studies could attempt to address such limitations by attempting to identify the recruiters responsible for making interviewing decisions.

Our gender analyses were limited by the fact that few resumes with male names were sent in response to advertisements for administrative positions. Consequently, although the impact of resume quality on callbacks appeared to vary by gender, we could not rule out the possibility that this effect was driven by differences in sales and administrative positions, thus limiting our ability to draw strong theoretical inferences from this finding. This finding, however, provides strong motivation for future research examining the interactive effects of resume characteristics and applicant gender in a variety of contexts. Specifically, future research should investigate the differential effects of application quality by gender within employment contexts where different genders are traditionally overrepresented as well as contexts where gender representation tends to more accurately reflect population demographics.

One final direction for future research would be related to the alphabetical sorting of applicant last names. Alphabetization of resumes can bias selection processes so strongly that one employee responsible for recruiting interns at an investment bank reported assigning 80 percent of interview slots halfway through an alphabetized file of 240 resumes (Yun, 2021). In other words, applicants in the latter half of the alphabet were four times less likely to receive an invitation to interview than those in the beginning of the process. Alphabetically sorting candidates is such a pervasive procedure in human resource management that Indeed (2022) lists this as a product benefit. B&M's dataset did not include last names and these data were not available from the authors (M. Bertrand, personal communication, March 8, 2022) so we were unable to directly explore the relationship between candidate last names and the likelihood of receiving a callback. We did, however, test this using a simulated value for last name sorting (see the OSF documents for more details) and found no evidence that the alphabetic order of last names influenced the likelihood of receiving a callback. This potential relationship, however, should be explored more extensively in future research.

Conclusion

Bertrand and Mullainathan's (2004) paper provided clear evidence of discrimination in the job market during a modern time period when many people wanted to believe that such behavior no longer existed. As a result, the paper became widely known, amassing more than 6,600 citations on route to becoming a seminal piece in the discrimination literature. The current research demonstrates that there is even more that can be gleaned from this study. While the current research may be disheartening in that it finds evidence of a racial hierarchy in terms of discrimination, it may also be encouraging in that it finds evidence of conditions in which human resource practices appear to reduce discriminatory effects. In a time where diversity scholars are challenged by the impact of social desirability on research participants, it is our hope that the current research will inspire scholars to look for additional insights that can be gleaned from prior work.

REFERENCES

ABC News. 2004, August 20. Can a "Black" name affect job prospects? *ABC News*. Retrieved March 2, 2016, from http://abcnews.go.com/2020/story?id=124232&page=1

ACENET. 2022. The Carnegie Classification of Institutions of Higher Education. *American Council on Education*. Retrieved June 1, 2022, from https://carnegieclassifications.acenet.edu/lookup/srp.php?clq=%7B%22basic2005_ids%22%3A%2215%22%7D&start_page=standard.php&backurl=standard.php&limit=0,50

Ai, C., & Norton, E. C. 2003. Interaction terms in logit and probit models. *Economic Letters*, 80(1): 123–129. https://doi.org/10.1016 /S0165-1765(03)00032-6.

Akala, A. 2020, September 23. Cost of racism: U.S. economy lost $16 trillion because of discrimination, bank says. *NPR*. Retrieved July 12, 2022, from https://www.npr.org/sections/live-updates-protests-for-racial-justice/2020/09/23/916022472/cost-of-racism-u-s-economy-lost-16-trillion-because-of-discrimination-bank-says

Anderson, J. 1991. Blame the Arabs: Tensions in the Gulf bring bigotry at home. *The Progressive*, 55(2): 28–29.

Arai, M., Bursell, M., & Nekby, L. 2016. The reverse gender gap in ethnic discrimination: Employer stereotypes of men and women with Arabic names. *The International Migration Review*, 50(2): 385–412. https://doi.org/10.1111/imre.l2170.

Asare, J. G. 2020, February 20. Are job candidates still being penalized for having 'ghetto ' names? *Forbes*. Retrieved August 10, 2022, from https://www.forbes.com/sites/janicegassam/2020/02/20/are-job-candidates-still-being-penalized-for-having-ghetto-names/?sh=9c2367650ed7

Åslund, O., & Skans, O. N. 2012. Do anonyomous job application procedures level the playing field? *ILR Review*, 65(1): 82–107. Retrieved from https://www.jstor.org/stable/41343666

Awad, G. H. 2010. The impact of acculturation and religious identification on perceived discrimination for Arab/Middle Eastern Americans. *Cultural diversity & ethnic minority psychology*, 16(1): 59–67. https://doi.org/10.1037/a0016675.

Benson, A., Li, D., & Shue, K. 2022. *"Potential" and the gender promotion gap*. Retrieved from https://danielle-li.github.io/assets/docs/PotentialAndTheGenderPromotionGap.pdf

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

Bertrand, M., & Mullainathan, S. 2004b. *Replication data for: Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination: lakisha_aer.dta*. Nashville, TN: Nashville, TN: American Economic Association [publisher]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-12-06. https://doi.org/10.3886/E116023V1.

Biernat, M., & Fuegen, K. 2001. Shifting standards and the evaluation of competence:

Complexity in gender-based judgment and decision making. *Journal of Social Issues*, 57(4): 707–724. https://doi.org/10.1111/0022-4537.00237.

Biernat, M., & Kobrynowicz, D. 1997. Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personality and Social Psychology*, 72(3): 544–557. https://doi.org/10.1037/0022-3514.72.3.544.

Biernat, M., & Thompson, E. R. 2002. Shifting standards and contextual variation in stereotyping. *European Review of Social Psychology*, 12(1): 103–137. https://doi.org/10.1080/14792772143000030.

Biernat, M., & Vescio, T. K. 2002. She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behavior. *Personality and Social Psychology Bulletin*, 28(1): 66–77. https://doi.org/10.1177/0146167202281006.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. 2009. Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, 94(3): 567–582. https://doi.org/10.1037/a0014665.

Boldry, J., Wood, W., & Kashy, D. A. 2001. Gender stereotypes and the evaluation of men and women in military training. *Journal of Social Issues*, 57(4): 689–705. https://doi.org/10.1111/0022-4537.00236.

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

Bushman, B. J., & Bonacci, A. M. 2004. You've got mail: Using e-mail to examine the effect of prejudiced attitudes on discrimination against Arabs. *Journal of Experimental Social Psychology*, 40(6): 753–759. https://doi.org/10.1016/j.jesp.2004.02.001.

Cambridge University Press. 2016. Familiar. *Cambridge Dictionaries Online*. Retrieved October 12, 2017, from http://dictionary.cambridge.org/us/dictionary/english/familiar

Castilla, E. J. 2015. Accounting for the gap: A firm study manipulating organizational accountability and transparency in pay decisions. *Organization Science*, 26(2): 311–333. https://doi.org/10.1287/orsc.2014.0950.

Chaffin, J. 2022, February 1. NFL sued for racial discrimination by ex-Miami Dolphins coach. *The Financial Times*. Retrieved July 17, 2022, from https://www.ft.com/content/4a83c01b-7db4-4603-80b7-39c15d85e95f

Charles, K. K., & Guryan, J. 2008. Prejudice and wages: An empirical assessment of Becker's The Economics of Discrimination. *Journal of Political Economy*, 116(5): 773–809. https://doi.org/10.1086/593073.

Chatla, S., & Shmueli, G. 2016. *Linear probability models (LPM) and big data : The good, the bad, and the ugly*. *Indian School of Business Research Paper Series*. Retrieved from http://ssrn.com/abstract=2353841

Cortina, L. M., Kabat-Farr, D., Leskinen, E. A., Huerta, M., & Magley, V. J. 2013. Selective incivility as modern discrimination in organizations: evidence and impact. *Journal of Management*, 39(6): 1579–1605. https://doi.org/10.1177/0149206311418835.

Costello, D. 2022, July 15. U.S. Department of Justice opens probe into Maryland State Police over possible racial discrimination in hiring and promotions. *The Baltimore Sun*. Retrieved July 17, 2022, from https://www.baltimoresun.com/news/crime/bs-md-cr-maryland-state-police-department-of-justice-20220715-wb3fluzekzdyzlfk4bq3a7fppm-story.html

Cotter, D. A. ., Hermsen, J. M. ., Ovadia, S., & Vanneman, R. 2001. The glass ceiling effect. *Social Forces*, 80(2): 655–681. https://doi.org/10.1353/sof.2001.0091.

Cotton, J. L., O'Neill, B. S., & Griffin, A. 2008. The "Name Game": Affective and hiring reactions to first names. *Journal of Managerial Psychology*, 23(1): 18–39. https://doi.org/10.1108/02683940810849648.

Daniels, S., & Thornton, L. D. M. 2019. Race and workplace discrimination: The mediating role of cyber incivility and interpersonal incivility. *Equality, Diversity and Inclusion*, 39(3): 319–335. https://doi.org/10.1108/EDI-06-2018-0105.

Derous, E., Nguyen, H.-H., & Ryan, A. M. 2009. Hiring discrimination against Arab minorities: Interactions between prejudice and job characteristics. *Human Performance*, 22(4): 297–320. https://doi.org/10.1080/08959280903120261.

Derous, E., Ryan, A. M., & Nguyen, H.-H. D. 2012. Multiple categorization in resume screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings. *Journal of Organizational Behavior*, 33(4): 544–570. https://doi.org/10.1002/job.769.

Dovidio, J. F., & Gaertner, S. L. 2000. Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11(4): 315–319. Retrieved from https://www.jstor.org/stable/40063839

DPH. 1979. Birth name count. State of Massachusetts Department of Public Health.

Drayton, N. 2013, September 18. Will a 'Black' name brand my son with mug shots before he's even born? *The New York Times*. Retrieved November 1, 2017, from http://parenting.blogs.nytimes.com/2013/09/18/will-a-black-name-brand-my-son-with-mug-shots-before-hes-even-born/

Duffy, R. 2016, September 15. 98FM accused of "racial discrimination" in job application by Dublin woman. *Yahoo! News*. Retrieved November 1, 2017, from https://uk.news.yahoo.com/98fm-accused-racial-discrimination-job-application-dublin-woman-144843148.html

Erwin, P. G. 1993. First names and perceptions of physical attractiveness. *Journal of Psychology*, 127(6): 625–631. https://doi.org/10.1080/00223980.1993.9914901.

Foley, M., & Williamson, S. 2018. Does anonymising job applications reduce gender bias?: Understanding managers' perspectives. *Gender in Management*, 33(8): 623–635. https://doi.org/10.1108/GM-03-2018-0037.

Friedman, S., Squires, G. D., & Chadwick, J. 2006. Cybersegregation: Are Neil Kelly and Greg Baker more desirable tenants than Tyrone Jackson or Jorge Rodriguez? A research proposal. *SAGE Race Relations Abstracts*, 31(2): 26–31. https://doi.org/10.1177/0307920106063069.

Fussell, S. 2016, May 13. Tinder for jobs aims to shatter hiring barriers in the tech world. *The Sidney Morning Herald*. Retrieved November 12, 2017, from http://www.smh.com.au/business/workplace-relations/tinder-for-jobs-aims-to-shatter-hiring-barriers-in-the-tech-world-20160513-gou80p.html

Gaddis, S. M., & Ghoshal, R. 2015. Arab American housing discrimination, ethnic competition, and the contact hypothesis. *The Annals of the American Academy of Political and Social Science*, 660(1): 282–299. https://doi.org/10.1177/0002716215580095.

Goodman, J. S., Fields, D. L., & Blum, T. C. 2003. Cracks in the glass ceiling. *Group and Organization Management*, 28(4): 475–501. https://doi.org/10.1177/1059601103251232.

Gündemir, S., Carton, A. M., & Homan, A. C. 2019. The impact of organizational performance on the emergence of Asian American leaders. *Journal of Applied Psychology*, 104(1): 107–122. https://doi.org/10.1037/apl0000347.

Hanks, P., Hardcastle, K., & Hodges, F. 2006. *A dictionary of first names* (Second edi.). New York: Oxford University Press.

Hanson, A., & Hawley, Z. 2011. Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in U.S. cities. *Journal of Urban Economics*, 70(2–3): 99–114. https://doi.org/10.1016/j.jue.2011.02.003.

Haslam, S. A., & Ryan, M. K. 2008. The road to the glass cliff: Differences in the perceived suitability of men and women for leadership positions in succeeding and failing organizations. *Leadership Quarterly*, 19(5): 530–546. https://doi.org/10.1016/j.leaqua.2008.07.011.

Herr, P. M. 1986. Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology*, 51(6): 1106–1115. https://doi.org/10.1037/0022-3514.51.6.1106.

Hilliard, J. 2022, February 13. Newton settles with firefighters: Two alleged enduring racial discrimination while in the workplace. *Boston Globe*. Retrieved July 17, 2022, from https://www.bostonglobe.com/2022/02/09/metro/newton-approves-two-settlements-firefighters-who-alleged-discrimination-job/

Indeed. 2022. Managing candidates. *Employer Help Center*. Retrieved May 15, 2022, from https://indeed.force.com/employerSupport1/s/article/360026204071?language=en_US

Jalal, S. 2016, September 14. @Sammm_J. *Twitter*. Retrieved November 12, 2017, from https://twitter.com/Sammm_J/status/776097534930477056

Judge, T. A., & Livingston, B. A. 2008. Is the gap more than gender? A longitudinal analysis of gender, gender role orientation, and earnings. *Journal of Applied Psychology*, 93(5): 994–1012. https://doi.org/10.1037/0021-9010.93.5.994.

Kaikati, J. G. 1978. The Arab boycott: Middle East business dilemma. *California Management Review*, 20(3): 32–46. https://doi.org/10.2307/41165280.

Kellar, S. J., & Hall, E. V. 2022. Measuring racial discrimination remotely: A contemporary review of unobtrusive measures. *Perspectives on Psychological Science*, 17(5): 1404–1430. https://doi.org/10.1177/17456916211045691.

King, E. B., & Ahmad, A. S. 2010. An experimental field study of interpersonal discrimination toward Muslim job applicants. *Personnel Psychology*, 63(4): 881–906. https://doi.org/10.1111/j.1744-6570.2010.01199.x.

Koch, A. J., D'Mello, S. D., & Sackett, P. R. 2015. A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, 100(1): 128–161. https://doi.org/10.1037/a0036734.

Köhler, T., & Cortina, J. M. 2021. Play it again, Sam! An analysis of constructive replication in the organizational sciences. *Journal of Management*, 47(2): 488–518. https://doi.org/10.1177/0149206319843985.

Kübler, D., Schmid, J., & Stüber, R. 2018. Gender discrimination in hiring across occupations : a nationally-representative vignette study. *Labour Economics*, 55: 215–229. https://doi.org/10.1016/j.labeco.2018.10.002.

Lebuda, I., & Karwowski, M. 2013. Tell me your name and I'll tell you how creative your work is: Author's name and gender as factors influencing assessment of products' creativity in four different domains. *Creativity Research Journal*, 25(1): 137–142. https://doi.org/10.1080/10400419.2013.752297.

Lee, S. Y., Pitesa, M., Thau, S., & Pillutla, M. M. 2015. Discrimination in selection decisions: Integrating stereotype fit and interdependence theories. *Academy of Management Journal*, 58(3): 789–812. https://doi.org/10.5465/amj.2013.0571.

Lerner, S. 1986, July 28. Terror against Arabs in America: No more looking the other way. *The New Republic*, 20–25.

Lewis, A. C., & Sherman, S. J. 2003. Hiring you makes me look bad: Social-identity based reversals of the ingroup favortism effect. *Organizational Behavior and Human Decision Processes*, 90(2): 262–276. https://doi.org/10.1016/S0749-5978(02)00538-1.

Lord, R. G., & Maher, K. J. 1993. *Leadership and information processing: Linking perceptions and performance*. New York: Routledge.

Lord, R. G., & Smith, J. E. 1983. Theoretical, information processing, and situational factors affecting attribution theory models of organizational behavior. *Academy of Management Review*, 8(1): 50–60. https://doi.org/10.5465/AMR.1983.4287658.

McWilliams, A., & Siegel, D. 1997. Event studies in management research: Theoretical and empirical issues. *Academy of Management Journal*, 40(3): 626–657. https://doi.org/10.2307/257056.

Mehrabian, A. 2001. Characteristics attributed to individuals on the basis of their first names. *Genetic, Social, and General Psychology Monographs*, 127(1): 59–88.

Mehrabian, A., & Piercy, M. 2001. Differences in positive and negative connotations of nicknames and given names. *The Journal of Social Psychology*, 133(5): 737–739. https://doi.org/10.1080/00224545.1993.9713930.

Milkman, K. L., Akinola, M., & Chugh, D. 2015. What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, 100(6): 1678–1712.

https://doi.org/10.1037/apl0000022.

Neumark, D., Bank, R. J., & Nort, K. D. Van. 1996. Sex discrimination in restaurant hiring: An audit study. *Quarterly Journal of Economics*, 111(3): 915–941. https://doi.org/10.2307/2946676.

Newton, D., & Simutin, M. 2015. Of age, sex, and money: Insights from corporate officer compensation on the wage inequality between genders. *Management Science*, 61(10): 2355–2375. https://doi.org/10.1287/mnsc.2014.1998.

Nkomo, S. M., Bell, M. P., Roberts, L. M., Joshi, A., & Thatcher, S. M. B. 2019. Diversity at a critical juncture: New theories for a complex phenomenon. *Academy of Management Review*, 44(3): 498–517. https://doi.org/10.5465/amr.2019.0103.

O'Connor, S. 2016, October 17. Q&A: does "name-blind" recruitment combat bias? *Financial Times*. Retrieved November 1, 2017, from https://www.ft.com/content/3d7b6590-9443-11e6-a80e-bcd69f323a8b

Obenauer, W. G., & Kalsher, M. J. 2022. Is white always the standard? Using replication to revisit and extend what we know about the leadership prototype. *The Leadership Quarterly*, 101633. https://doi.org/10.1016/j.leaqua.2022.101633.

Obenauer, W. G., & Langer, N. 2019. Inclusion is not a slam dunk: A study of differential leadership outcomes in the absence of a glass cliff. *Leadership Quarterly*, 30(6): 101334. https://doi.org/10.1016/j.leaqua.2019.101334.

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

Ozturk, M. B., & Berber, A. 2022. Racialised professionals' experiences of selective incivility in organisations: A multi-level analysis of subtle racism. *Human Relations*, 75(2): 213–239. https://doi.org/10.1177/0018726720957727.

Pager, D., & Quillian, L. 2005. Walking the talk? What employers say versus what they do. *American Sociological Review*, 70(3): 355–380. https://doi.org/10.1177/000312240507000301.

Park, S. H., & Westphal, J. D. 2013. Social discrimination in the corporate elite: How status affects the propensity for minority CEOs to receive blame for low firm performance. *Administrative Science Quarterly*, 58(4): 542–586. https://doi.org/10.1177/0001839213509364.

Parkinson, J., & Smith-Walters, M. 2015, October 25. Who, what, why: What is name-blind recruitment? *BBC News Magazine*. Retrieved November 1, 2017, from http://www.bbc.com/news/magazine-34636464

Pratto, F., Stallworth, L. M., Sidanius, J., & Siers, B. 1997. The gender gap in occupational role attainment: A social dominance approach. *Journal of Personality & Social Psychology*, 72(1): 37–53. https://doi.org/10.1037/0022-3514.72.1.37.

Prime Minister's Office. 2015, October 26. PM: Time to end discrimination and finish the fight for real equality. *gov.uk*. Retrieved November 1, 2017, from https://www.gov.uk/government/news/pm-time-to-end-discrimination-and-finish-the-fight-for-real-equality

Rooth, D.-O. 2010. Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3): 523–534. https://doi.org/10.1016/j.labeco.2009.04.005.

Rosch, E. 1978. Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48).

Rosette, A. S., Carton, A. M., Bowes-Sperry, L., & Hewlin, P. F. 2013. Why do racial slurs remain prevalent in the workplace? Integrating theory on intergroup behavior. *Organization Science*, 24(5): 1402–1421. https://doi.org/10.1287/orsc.1120.0809.

Rosette, A. S., Leonardelli, G. J., & Phillips, K. W. 2008. The white standard: Racial bias in leader categorization. *Journal of Applied Psychology*, 93(4): 758–777. https://doi.org/10.1037/0021-9010.93.4.758.

Ryan, M. K., & Haslam, S. A. 2005. The glass cliff: Evidence that women are over-represented in precarious leadership positions. *British Journal of Management*, 16(2): 81–90. https://doi.org/10.1111/j.1467-8551.2005.00433.x.

Ryan, M. K., Haslam, S. A., & Postmes, T. 2007. Reactions to the glass cliff: Gender differences in the explanations for the precariousness of women's leadership positions. *Journal of Organizational Change Management*, 20(2): 182–197. https://doi.org/10.1108/09534810710724748.

Saad, L. 2006, August 10. Anti-Muslim sentiments fairly commonplace: Four in ten Americans admit feeling prejudice against Muslims. *Gallup News Service*. Retrieved November 1, 2017, from http://media.gallup.com/worldpoll/pdf/antimuslimsentiment81006.pdf

Salerno, J. M., Peter-Hagene, L. C., & Jay, A. C. V. 2019. Women and African Americans are less influential when they express anger during group decision making. *Group Processes and Intergroup Relations*, 22(1): 57–79. https://doi.org/10.1177/1368430217702967.

Semaan, G. 2014. Arab Americans: Stereotypes, conflict, history, cultural identity and post 9/11. *Intercultural Communication Studies*, 23(2): 17–32. Retrieved from https://www-s3-live.kent.edu/s3fs-root/s3fs-public/file/Gaby-Semaan.pdf

Shaheen, J. G. 2003. Reel bad Arabs: How Hollywood vilifies a people. *The Annals of the American Academy of Political and Social Science*, 588(1): 171–193. Retrieved from https://www.jstor.org/stable/1049860

Srull, T. K., & Wyer, R. S. 1979. The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, 37(10): 1660–1672. https://doi.org/10.1037/0022-3514.37.10.1660.

Stauffer, J. M., & Buckley, M. R. 2005. The existence and nature of racial bias in supervisory ratings. *Journal of Applied Psychology*, 90(3): 586–591. https://doi.org/10.1037/0021-9010.90.3.586.

Stockton, R. 1994. Ethnic archetypes and the Arab Image. In E. McCarus (Ed.), *The Development of Arab-American Identity* (pp. 119–153). University of Michigan Press.

Stone-Romero, E. F., & Anderson, L. E. 1994. Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal of Applied Psychology*, 79(3): 354–359. https://doi.org/10.1037/0021-9010.79.3.354.

Suleiman, M. W. 1999. Islam, Muslims and Arabs in America: the other of the other of the other ... *Journal of Muslim Minority Affairs*, 19(1): 33–47. https://doi.org/10.1080/13602009908716423.

Tadmor, C. T., Hong, Y., Chao, M. M., Wiruchnipawan, F., & Wang, W. 2012. Multicultural experiences reduce intergroup bias through epistemic unfreezing. *Journal of Personality and Social Psychology*, 103(5): 750–72. https://doi.org/10.1037/a0029719.

U.S. Equal Employment Opportunity Commission. 2023. Newsroom. Retrieved January 15, 2023, from https://www.eeoc.gov/newsroom

Ubaka, A., Lu, X., & Gutierrez, L. 2022. Testing the generalizability of the white leadership standard in the post-Obama era. *Leadership Quarterly*, 101591. https://doi.org/10.1016/j.leaqua.2021.101591.

Uhlmann, E. L., & Cohen, G. L. 2007. "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, 104(2): 207–223. https://doi.org/10.1016/j.obhdp.2007.07.001.

Umphress, E. E., Simmons, A. L., Boswell, W. R., & Triana, M. d. C. 2008. Managing discrimination in selection: the influence of directives from an authority and social dominance orientation. *Journal of Applied Psychology*, 93(5): 982–993. https://doi.org/10.1037/0021-9010.93.5.982.

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

Wilcox, A., Damarin, A. K., & McDonald, S. 2022. Is cybervetting valuable? *Industrial and Organizational Psychology*, 15(3): 315–333. https://doi.org/10.1017/iop.2022.28.

Wright, E. O., & Baxter, J. 2000. The glass ceiling hypothesis : A reply to critics. *Gender and Society*, 14(6): 814–821. Retrieved from https://www.jstor.org/stable/190377

Young, R. K., Kennedy, A. H., Newhouse, A., Browne, P., & Thiessen, D. 1993. The effects of names on perception of intelligence, popularity, and competence. *Journal of Applied Social Psychology*, 23(21): 1770–1788. https://doi.org/10.1111/j.1559-1816.1993.tb01065.x.

Yun, W. 2021, October 18. Op-Ed : Mind your XYZs for those of us with names at the back of the line. *Los Angeles Times*. Retrieved June 1, 2022, from https://www.latimes.com/opinion/story/2021-10-18/op-ed-mind-your-xyzs-for-those-of-us-with-names-at-the-back-of-the-line

Zapata, C. P., Carton, A. M., & Liu, J. T. 2016. When justice promotes injustice: Why minority leaders experience bias when they adhere to interpersonal justice rules. *Academy of Management Journal*, 59(4): 1150–1173. https://doi.org/10.5465/amj.2014.0275.

TABLE 1

Mean Callback Rates by Condition

Replication Type	Model	Sample	Manipulation	Cond 1	Cond 2	Ratio	Diff (SE)	z-stat (p-value)
Direct	A	All	White (1) vs. Not White (2)	9.651% [2,435]	6.448% [2,435]	1.497	3.203% (0.008)	4.110 (.000)
Direct	B	Chicago	White (1) vs. Not White (2)	8.062% [1,352]	5.399% [1,352]	1.493	2.663% (0.010)	2.763 (.006)
Direct	C	Boston	White (1) vs. Not White (2)	11.634% [1,083]	7.756% [1,083]	1.500	3.878% (0.013)	3.050 (.002)
Direct	D	Females	White (1) vs. Not White (2)	9.889% [1,860]	6.628% [1,886]	1.492	3.261% (0.009)	3.630 (.000)
Direct	E	Females in Admin Jobs	White (1) vs. Not White (2)	10.457% [1,358]	6.549% [1,359]	1.597	3.908% (0.011)	3.650 (.000)
Direct	F	Females in Sales Job	White (1) vs. Not White (2)	8.367% [502]	6.831% [507]	1.225	1.535% (0.017)	0.930 (.353)
Direct	G	Males	White (1) vs. Not White (2)	8.870% [575]	5.829% [549]	1.522	3.041% (0.009)	1.950 (.051)
Direct	H	White	High Quality (1) vs. Low Quality (2)	10.793% [1223]	8.498% [1212]	1.270	2.295% (0.012)	1.917 (.055)
Direct	I	Not-White	High Quality (1) vs. Low Quality (2)	6.705% [1223]	6.188% [1212]	1.083	0.517% (0.010)	0.519 (.604)
Conceptual	J	High Quality Resume	White (1) vs. Not White (2)	10.793% [1,223]	6.705% [1,223]	1.610	4.088% (0.011)	3.580 (.000)
Conceptual	K	Low Quality Resume	White (1) vs. Not White (2)	8.498% [1,212]	6.188% [1,212]	1.373	2.310% (0.011)	2.180 (.029)
Conceptual	L	Requirements = No	White (1) vs. Not White (2)	13.127% [518]	7.336% [518]	1.789	5.792% (0.019)	3.080 (.002)
Conceptual	M	Requirements = Yes	White (1) vs. Not White (2)	8.712% [1917]	6.208% [1917]	1.403	2.504% (0.008)	2.950 (.003)
Conceptual	N	White	No Job Req (1) Job Req (2)	13.127% [518]	8.712% [1917]	1.507	4.416% (0.012)	3.020 (.003)
Conceptual	O	Not White	No Job Req (1) Job Req (2)	7.336% [518]	6.208% [1917]	1.182	1.128% (0.011)	0.928 (.354)
Conceptual	P	Not White	Not Arabic (1) vs. Arabic (2)	7.022% [2,008]	3.747% [427]	1.874	3.275% (0.011)	2.502 (.012)
Conceptual	Q	All	White (1) vs. Arabic (2)	9.651% [2,435]	3.747% [427]	2.576	5.904% (0.011)	3.978 (.000)
Conceptual	R	Matched Comparisons	White (1) vs. Arabic (2)	8.665% [427]	3.747% [427]	2.313	4.918% (0.016)	2.979 (.003)
Conceptual	S	Female	High Quality (1) vs. Low Quality (2)	9.207% [1,879]	7.284% [1,867]	1.264	1.923% (0.009)	2.139 (.033)
Conceptual	T	Male	High Quality (1) vs. Low Quality (2)	7.231% [567]	7.540% [557]	0.959	-0.309% (0.016)	0.198 (.843)
Conceptual	U	Match Female	High Quality (1) vs. Low Quality (2)	9.290% [1,507]	6.835% [1,507]	1.359	2.455% (0.010)	2.475 (.013)
Conceptual	V	Match Male	High Quality (1) vs. Low Quality (2)	7.510% [253]	7.905% [253]	0.950	-0.395% (0.024)	0.167 (.868)
Conceptual	W	Match Female Sales	High Quality (1) vs. Low Quality (2)	6.373% [204]	5.392% [204]	1.182	0.980% (0.023)	0.421 (.674)
Conceptual	X	Match Male Sales	High Quality (1) vs. Low Quality (2)	7.111% [225]	7.556% [225]	0.941	-0.444% (0.025)	0.181 (.857)

Note: Table reports the callback rates for each condition defined in the manipulation column, the sample size per condition in [brackets], the ratio and difference of callback rates, along with the standard error of the difference and p-values reported in tests of proportion.

TABLE 2

Probit Regressions Estimating Effects of Demographic Traits and Resume Characteristics on Likelihood of Callback

Model	A	B	C	D	E	F	G	H	I	J	K	L
Sample	All	All	All	White	Not White	All[a]	No Exp Req[a]	Exp Req[a]	All	All	All	All
Applicant Race (NW=1)	-0.025 [.004] (0.008)	N/A	-0.031 [.000] (0.006)	N/A	N/A	0.021 [.485] (0.030)	0.044 [.356] (0.047)	-0.014 [.723] (0.040)	-0.026 [.000] (0.007)	-0.023 [.232] (0.020)	N/A	-0.032 [.000] (0.006)
Years of Experience	N/A	0.007 [.010] (0.003)	0.007 [.010] (0.003)	0.014 [.001] (0.004)	0.002 [.496] (0.003)	-0.010 **[.036]** (0.005)	-0.014 **[.027]** (0.007)	-0.002 [.797] (0.008)	0.007 [.010] (0.003)	0.007 [.010] (0.003)	0.007 [.010] (0.003)	N/A
Experience Squared	N/A	0.000 [.049] (0.000)	0.000 [.049] (0.000)	0.000 [.004] (0.000)	0.000 [.949] (0.000)	0.000 **[.030]** (0.000)	0.001 **[.012]** (0.000)	0.000 [.830] (0.000)	0.000 [.047] (0.000)	0.000 [.047] (0.000)	0.000 [.047] (0.000)	N/A
Volunteer Experience	N/A	-0.006 [.588] (0.011)	-0.005 [.642] (0.011)	-0.016 [.311] (0.016)	0.007 [.643] (0.014)	0.022 [.324] (0.022)	0.077 [.068] (0.042)	-0.020 [.374] (0.023)	-0.005 [.639] (0.011)	-0.005 [.641] (0.011)	-0.005 [.650] (0.011)	N/A
Military Experience	N/A	0.001 [.957] (0.015)	0.002 [.919] (0.015)	0.023 [.368] (0.026)	-0.016 [.350] (0.017)	-0.032 [.155] (0.023)	-0.021 [.538] (0.035)	-0.024 [.525] (0.038)	0.002 [.912] (0.015)	0.002 [.913] (0.015)	0.001 [.921] (0.015)	N/A
Email	N/A	0.016 [.152] (0.011)	0.015 [.182] (0.011)	0.029 [.074] (0.016)	0.000 [.997] (0.013)	-0.024 [.158] (0.017)	-0.060 **[.008]** (0.023)	0.002 [.927] (0.024)	0.014 [.191] (0.011)	0.014 [.192] (0.011)	0.014 [.193] (0.011)	N/A
Employment Holes	N/A	0.024 [.029] (0.011)	0.024 [.029] (0.011)	0.034 [.038] (0.016)	0.014 [.303] (0.013)	-0.015 **[.365]** (0.017)	-0.028 [.178] (0.021)	0.000 [.998] (0.028)	0.024 [.028] (0.011)	0.024 [.028] (0.011)	0.024 [.028] (0.011)	N/A
Worked While in School	N/A	0.010 [.284] (0.009)	0.009 [.313] (0.009)	0.020 [.157] (0.014)	-0.001 [.920] (0.012)	-0.017 [.303] (0.017)	-0.023 [.321] (0.023)	-0.017 [.458] (0.023)	0.009 [.311] (0.009)	0.009 [.312] (0.009)	0.009 [.311] (0.009)	N/A
Honors Listed	N/A	0.049 [.019] (0.021)	0.049 [.019] (0.021)	0.063 [.058] (0.033)	0.031 [.205] (0.025)	-0.012 [.659] (0.027)	-0.031 [.280] (0.028)	0.020 [.731] (0.057)	0.048 [.020] (0.021)	0.048 [.020] (0.021)	0.048 [.020] (0.021)	N/A
Computer Skills	N/A	-0.022 [.089] (0.013)	-0.019 [.119] (0.012)	-0.043 [.026] (0.019)	-0.001 [.932] (0.014)	0.029 **[.154]** (0.021)	0.030 [.291] (0.029)	0.031 [.294] (0.029)	-0.020 [.115] (0.012)	-0.020 [.115] (0.012)	-0.020 [.113] (0.012)	N/A

TABLE 2 (CONTINUED)

Probit Regressions Estimating Effects of Demographic Traits and Resume Characteristics on Likelihood of Callback

Model	A	B	C	D	E	F	G	H	I	J	K	L
Sample	All	All	All	White	Not White	All[a]	No Exp Req[a]	Exp Req[a]	All	All	All	All
Special Skills	N/A	0.055 [.000] (0.012)	0.055 [.000] (0.012)	0.066 [.000] (0.016)	0.043 [.001] (0.013)	-0.010 **[.448]** (0.013)	-0.008 **[.668]** (0.019)	-0.009 **[.620]** (0.018)	0.055 [.000] (0.012)	0.055 [.000] (0.012)	0.054 [.000] (0.012)	N/A
Female	N/A	0.002 [.827] (0.011)	0.003 [.769] (0.011)	0.002 [.915] (0.017)	0.005 [.715] (0.013)	N/A	N/A	N/A	-0.001 [.923] (0.011)	-0.001 [.910] (0.012)	-0.003 [.780] (0.012)	-0.003 [.858] (0.015)
Resume Quality	0.020 [.012] (0.008)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	-0.004 [.775] (0.014)
Resume Quality * Race	-0.013 [.204] (0.011)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Resume Quality * Female	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.024 [.184] (0.018)
Arabic Name	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	-0.029 [.027] (0.013)	-0.029 [.028] (0.013)	-0.030 [.018] (0.013)	N/A
Name Frequency	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.001 [.893] (0.005)	0.005 [.001] (0.002)	N/A
Wald Chi-Squared	32.128	83.25214	99.400	83.38699	41.1455	125.728	130.871	66.65815	99.094	99.169	96.331	32.839
p-value	.000	.000	.000	.000	.008	.000	.000	.000	.000	.000	.000	.000
SD (predict callback)	0.018	0.047	0.050	0.062	0.037	0.053	0.066	0.062	0.050	0.050	0.050	0.019
Sample Size	4,870	4,870	4,870	2,435	2,435	4,870	2,750	2,120	4,870	4,870	4,870	4,870

Notes: Table reports marginal effects, [p-values], and (standard errors of marginal effects)

[a]Models F, G, and H are interaction models. Reported values are for main effect of race and interaction of race with explanatory variables, bold font = direct effect (not shown) significant at p<0.05

TABLE 3

Population Frequency of Names Used

Female Names Used						Alternative Female Name Bank					
White			Black / African-American			White			Black / African-American		
Name	Name Count[a]	% of Births	Name	Name Count	% of Births	Name	Name Count	% of Births	Name	Name Count	% of Births
Sarah	3622	2.268%	Latoya	67	0.042%	Beverly	67	0.042%	Latoya[b]	67	0.042%
Jill	1518	0.951%	Aisha	53	0.033%	Madeline	55	0.034%	Aisha	53	0.033%
Allison	1404	0.879%	Tamika	41	0.026%	Sophia	43	0.027%	Tamika	41	0.026%
Emily	974	0.610%	Tanisha	33	0.021%	Alexa	33	0.021%	Tanisha	33	0.021%
Kristen	971	0.608%	Keisha	31	0.019%	Brandie	31	0.019%	Keisha	31	0.019%
Anne	878	0.550%	Ebony	27	0.017%	Belinda	27	0.017%	Ebony	27	0.017%
Carrie	737	0.462%	Kenya	26	0.016%	Athena	26	0.016%	Kenya	26	0.016%
Laurie	612	0.383%	Lakisha	25	0.016%	Marina	25	0.016%	Lakisha	25	0.016%
Meredith	441	0.276%	Latonya	5	0.003%	Larissa	21	0.013%	Nakia	21	0.013%

Male Names Used						Alternative Male Name Bank					
White			Black / African-American			White			Black / African-American		
Name	Name Count	% of Births	Name	Name Count	% of Births	Name	Name Count	% of Births	Name	Name Count	% of Births
Jay	7112	3.918%	Tyrone	86	0.047%	Dylan	63	0.035%	Jamal	63	0.035%
Matthew	6462	3.560%	Jamal	63	0.035%	Clayton	39	0.021%	Willie	38	0.021%
Geoffrey	3900	2.148%	Jermaine	34	0.019%	Jamison	34	0.019%	Jermaine	34	0.019%
Greg	1835	1.011%	Leroy	18	0.010%	Lewis	19	0.010%	Lamont	19	0.010%
Todd	1099	0.605%	Hakim	7	0.004%	Colby	18	0.010%	Leroy	18	0.010%
Brendan	489	0.269%	Rasheed	6	0.003%	Carlton	15	0.008%	Marvin	16	0.009%
Neil	410	0.226%	Darnell	6	0.003%	Abel	7	0.004%	Hakim	7	0.004%
Brad	366	0.202%	Tremayne	5	0.003%	Perry	6	0.003%	Clifton	6	0.003%
Brett	363	0.200%	Kareem	5	0.003%	Elvin	6	0.003%	Darnell	6	0.003%

		Name Ct	% Pop			Name Ct	% Pop
Mean	White Names Used	1844.056	1.063%	Mean	Alternative White Names	29.722	0.018%
	Black. Names Used	29.889	0.018%		Alternative Black. Names	29.500	0.018%

Notes: Names used refers to the actual first names used B&M's study. Alternative names refer to names that could have been used that would have met their basic requirements for "White" and "African-American" names but would have had similar frequency within the overall population. Name count represents the number of times that a first name appeared on birth certificates in the state of Massachusetts between 1974 and 1979. Percent of births refers to the percentage of all births in Massachusetts during that time period where a specific name was given to a child. Shaded areas were used to identify first names that appeared both in our alternative name bank and the original study

FIGURE I

95 Percent Confidence Intervals for Marginal Effects of Resume Characteristics by Applicant Race