A JSM approach to modern bias in employment interviews and possible remediation

Alexander Nassrelgrgawi

University at Albany, State University of New York, alexander9091@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Psychology Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
A JSM Approach to Modern Bias in Employment Interviews and Possible Remediation

by

Alexander Sharif Nassrelgrgawi

A Dissertation Submitted
to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy
College of Arts and Sciences
Department of Psychology

May 2021
ABSTRACT

In the last 10 years, the EEOC saw a 200% increase in the number of ethnic based discrimination charges involving individuals who were perceived to be of an Arab background (Equal Employment Opportunity Commission, 2017). However, little empirical work has been done to detect and amend this trend. To answer this gap, this investigation examined the expression of modern bias through the lens of the justification-suppression model (Crandall & Eshleman, 2003) in order to detect and explain possible discrimination Arabs may experience during employment interviews. This was done in two studies using participants recruited from Amazon’s Mechanical Turk. Study 1 had a final sample of 319 participants, while Study 2 had a final sample of 418 participants. Both studies utilized the same materials for pre-recorded interviews of a male answering questions at low, moderate, or high levels of interview performance/competence—a variable that was manipulated between-subjects. Ratings of interview performance were aggregated and evaluated for potential differences between the White and Arab conditions, although no significant effect was present. Study two empirically tested the impact of the two diversity trainings on justified and suppressed bias. In Study two, participants first completed an online diversity training (goal setting, perspective taking, or control – internet safety). Next, they had two weeks to practice and display the behaviors they learned in their training sessions. Although, no significant hypotheses were supported in regards to the effects of diversity training across interview performance conditions, results directionally suggest that a goal setting methodology may be more efficacious in reducing bias than a perspective taking methodology. Implications of these findings are discussed relative to assisting organizations in reducing ethnic bias in employment interview settings.
Table of Contents

Abstract........................................................................................................................................... ii

Chapter 1 ........................................................................................................................................ 1
  Introduction ..................................................................................................................................... 1

Chapter 2 ......................................................................................................................................... 9
  Stigmatized Individuals and the Various Expressions of Discrimination ......................................... 9
  Modern Bias in Interviews ............................................................................................................. 14
  Justified and Suppressed Biases .................................................................................................... 17
  Diversity Training: A Possible Remedy for Bias in Employment Interviews .................................... 21
  Perspective Taking ....................................................................................................................... 26
  Goal Setting ................................................................................................................................. 27
  Current Study ............................................................................................................................... 32

Chapter 3 ....................................................................................................................................... 34
  Study one Method ......................................................................................................................... 34
    Participants ................................................................................................................................. 34
    Procedure ................................................................................................................................. 35
    Measures ................................................................................................................................. 37
    Analysis Strategy ...................................................................................................................... 39
  Study one Results ....................................................................................................................... 40
  Study one Discussion .................................................................................................................... 43
  Study two Method ....................................................................................................................... 45
    Participants ................................................................................................................................. 45
    Procedure ................................................................................................................................. 46
    Materials .................................................................................................................................... 48
    Measures .................................................................................................................................... 49
    Analysis Strategy ...................................................................................................................... 50
  Study two Results ....................................................................................................................... 51
  Study two Discussion .................................................................................................................... 62

Chapter 4 ....................................................................................................................................... 65
  General Discussion ....................................................................................................................... 65
  Study Limitations ........................................................................................................................ 66
  Implications ................................................................................................................................... 68
  Theoretical Contributions and Future Directions ............................................................................ 70
  Practical Contributions and Future Directions ............................................................................... 72
  Conclusion .................................................................................................................................... 73

References ....................................................................................................................................... 74

Appendix A: Interview Questionnaire ............................................................................................... 88
  Interview procedure .................................................................................................................... 88
  Ethnicity (Manipulation check) .................................................................................................... 92
  Warmth and Competence Questions ............................................................................................ 92

Appendix B: Diversity Training Materials (Study 2) ...................................................................... 94
  Perspective taking Materials ........................................................................................................ 96
  Goal Setting Materials ................................................................................................................ 96
  Internet safety materials .............................................................................................................. 97
Appendix C: Materials Created for Ethnicity Manipulation .......................................................... 101
Appendix D: Ancillary Analyses Explored ..................................................................................... 103

Hypothesis 1 Re-test at Each Interview Question Level (Study 1 Data) ......................................... 106
Chapter 1

Introduction

Job interviews have become an ubiquitously ingrained ritual in the hiring process over the last century (Binet, 1911; Ryan and Ployhart, 2014). To find a selection system without an interview component is rare, and for most organizations, removing an interview component from their selection system is unthinkable (Huffcutt and Culbertson, 2013). Despite the popularity of interviews (Steiner, 2012), past research has exhibited many sources of implicit biases (e.g., ethnicity and race) that exist and can artificially alter ratings on interview scores (Koch, D’Mello, and Sackett, 2015; Purkiss, Perrewe, Gillespie, Mayes, and Ferris, 2006). These biasing effects are concerning for both organizations and stigmatized individuals. From the organizations’ side, biased interview evaluations may lead to needlessly underutilizing certain talent populations, legal repercussions, and impacted firm performance (Koch, D’Mello, and Sackett, 2015). From the point of view of stigmatized individuals, stigmatized applicants may experience unfair treatment and evaluation in interview settings, requiring them to work disproportionately harder to achieve gainful employment.

Although focus on racism in personnel selection is not a new topic, a focus on Arab minorities has been largely ignored by Industrial/Organizational Psychology journals. Rather, the overwhelming majority of research on intergroup relations and race-based differences in employment testing and hiring has focused on Black-White differences (Crosby, Bromley, and Saxe 1980; Dipboye and Colella, 2005). Despite this lack of focus, current estimates suggest three-four million people identify with an Arab ethnicity in 2018 (Gallup Research Center, 2017; Pew Research Center, 2018). Moreover, since the tragedy of September 11, 2001, a growing body of work suggests Arabs and Arab-Americans are leading targets of prejudice. Arabs, who
are assumed to be Muslim, may be seen as terrorists and religious fanatics (Fiske, Cuddy, and Glick, 2007; Zogby International, 2007) and are often stereotypically labeled as “cruel, deceitful, hot-tempered, and irrational” (American-Arab Anti-Discrimination Committee, 2008, p. 76). The effect of these views cannot be understated: people reported greater prejudice toward Arab Americans than toward African Americans, Asian Americans, or Hispanic Americans (Bushman and Bonacci, 2004), and a study by the Pew Research Center (2018) has confirmed that Americans’ views toward Muslims have been worsening. Yet, despite the evidence that prejudice toward Arabs in America is on the rise, the consideration of how these prejudicial views are expressed in employment selection settings, in this case interviews, is still a relatively new area of systematic research (King and Ahmad, 2010).

With the growth and proliferation of affirmative action plans (Harrison, Kravitz, Mayer, Leslie, and Lev-Arey, 2006), diversity trainings, and diversity education (Alhejji, Garavan, Carbery, O'Brien, and McGuire, 2016), it should be expected that the effects of stigmatized biases should be diminishing. Some studies provide some marginal evidence that this is the case. Recently, researchers compared meta-analytic mean effect size differences on interview scores between Whites and Blacks for pre-1996 and post-1996 studies and found that these estimates were significant for pre-1996 studies but were non-significant for the post-1996 group (Levashina, Hartwell, Morgeson, and Campion, 2014). Readers may feel inclined to interpret this finding as indicating that bias is no longer a problem in interview ratings. However, given that the EEOC in 2015 received 89,385 private sector charges of discrimination (Equal Employment Opportunity Commission, 2017) and that this is an increase compared to twenty years ago (see Figure 1), there seems to be a gap between science and practice. More specifically, in the last 10 years, the EEOC saw a 200% increase in the number of ethnic based discrimination charges
involving individuals who were perceived to be of an Arab background. These claims often involve the inclusion of a religious discrimination component towards being perceived as Muslim, which has risen over 250% in the past 15 years (Equal Employment Opportunity Commission, 2016a; Equal Employment Opportunity Commission, 2017).

Figure 1. EEOC Complaint frequencies over the last twenty years

Figure 1 shows the number of complaints the EEOC has had over the last twenty years, with breakdowns on categorizations complaints were filed under. Complaints related to Race, Sex, and National Origin over the last twenty years have remained static or have seen slight increases while the literature is claiming a diminishment of bias. The reason for this gap may be due to several different factors such as people becoming more likely to report discrimination (e.g., the MeToo movement), discrimination occurring less on the formal and overt levels and more at the interpersonal and subtle levels, and the expression of modern discrimination differing from traditional forms of discrimination (Dipboye and Johnson, 2013, Macan, 2009). Given the lack of clarity on modern interview bias, especially for Arab Americans who are often targeted in
the current socio-political climate but not systematically researched as victims of discrimination, there is an important need to both understand and limit the expression of modern bias in the selection process.

Fundamentally, the pattern of how racial biases might influence interview scores is still unclear, with prior studies revealing small and inconsistent effects between Whites and non-Whites (Arvey and Campion, 1982; Harris, 1989; Macan, 2009; Posthuma, Morgeson, and Campion, 2002; Schmitt, 1976). This has led many researchers to conclude that racial bias does not play a significant role in the interview process – something that may seem hard to agree with, especially for stigmatized individuals who report experiences of discrimination, even at the highest academic institutions. For example, as of 2019, Harvard University is a defendant in a lawsuit from a former African American employee who alleges multiple forms of discriminatory treatment involving denial of promotions and opportunities for flexible schedules granted to other co-workers, and being the recipient of acts of sabotage intended to inhibit her ability to do her job (Chaidez and Ryan, 2019). To this former Harvard employee, as well as the tens of thousands of employees who file race-based discrimination claims in the U.S. every year, it would seem that race-based discrimination in the workplace is alive and well.

Posthuma, Morgeson, and Campion (2002) suggested a few reasons for the perplexing research finding that race tends to not be a significant negative predictor of interview performance that include positive outcomes of equal opportunity legislation and education, or alternatively, a lack of methodological rigor in the studies evaluating bias in interviews (e.g., the Hawthorne effect). However, a recently published meta-analysis by Quillian, Pager, Hexel, and Midtboen (2017) challenges these past findings by suggesting racial discrimination in hiring is still a problem that has been unchanged over the last 25 years. Clearly, there is still an open
question on the state of affairs for sub-group differences in hiring, especially as related to Arabs and Arab Americans. In fact, Ruggs et al. (2013) found that only 19 articles were published in top IO Psychology journals related to applicant/employee race between 1990 and 2013. This is troublesome for the field and practice at large because the inconclusive, deficient literature on the subject leaves readers to draw their own conclusions about the state of affairs regarding racial bias in interview settings. This in turn, may lead to the furthering of discriminatory workplace artifacts and the justification of discriminatory thoughts and beliefs.

One possible explanation for the aforementioned inconclusive findings can be explained through a theoretical model that accounts for unique circumstances where interviewers may feel justified in their biased judgements, or else motivated to suppress their biases (Crandall and Eshleman, 2003). According to Crandall and Eshleman's (2003) Justification-Suppression Model (JSM) of prejudice, people manage their biases by filtering the expression of prejudice through processes of suppression and justification, depending on whether situational cues that are presented challenge or reinforce stereotypes. Other researchers have developed stereotype content models that outline factors that may activate (justify) or inhibit (suppress) an interviewer’s expression of prejudice (Cuddy, Fiske, and Glick, 2004; Fiske, Cuddy, Glick, and Xu, 2002). Previous studies finding inconsistent effects of subgroup differences on interview evaluations may not have accounted for such factors that are theorized to encourage or suppress discrimination, thereby confounding their results. One such ignored factor is competence, or interview performance. For instance, a minority interviewee may perform well by answering questions with the desired responses, which in turn would motivate the interviewer to suppress their prejudice and rate the applicant more fairly. Conversely, the an interviewee may be underperforming and struggling to answer the interviewer’s questions, which would provide
justification for the interviewer rating this minority applicant more harshly than they would a non-minority of the same competence level. The JSM (Crandall and Eshleman, 2003) provides support for a finer grained exploration of sub-group differences on varying degrees of interviewee competence to explore whether race-based discrimination in interview ratings are perhaps more likely under only particular conditions—particularly at lower levels of interviewee competence where discrimination may be seen as more justified. Should these hypothesized differences exist, the next logical question is “What steps can we take to reduce or remove these observed effects?”. Research in interview prejudice has been sparse as far as remedies for the issues uncovered (Dovidio and Gaertner, 2000; King and Ahmad, 2010; McConahay, 1986). Instead, the recommendations very frequently focus on what the victims of these discriminatory acts can do in order to avoid the prejudice. However, related research from the diversity training literature suggests two training types, perspective taking and goal setting, as strategies that may reduce the effects of justified prejudice (Bezrukova, Jehn, and Spell, 2012). One limit to these findings, though, is that the validity for behavior change is somewhat lackluster, as most studies in this area examine attitude change or self-reported behavior change as outcome variables to determine the relative success of these diversity training interventions (Holladay and Quinones, 2008; Lindsey et al., 2015). The current investigation contributes to the diversity training literature by providing a more objective criterion (ratings of interview competence) to evaluate training efficacy, as well as providing more clarity regarding which diversity training may be the most effective in reducing the effects of prejudicial ratings.

In short, this proposed investigation presents two studies that have been designed to provide updated evidence on the following two research questions:
1. How does applicant interview performance and modern racial bias interact to influence interview ratings?

2. Should bias exist, what training related strategies can organizations employ to facilitate more accurate interview ratings?

This investigation plans to examine the expression of modern bias through the lens of the justification-suppression model (Crandall and Eshleman, 2003) in order to detect and explain any possible discrimination Arabs may experience during employment interviews. Additionally, should findings in Study 1 point to a confirmation of modern bias expressed in interview scores, then Study 2 will attempt to evaluate strategies that might ameliorate the problem. Without accounting for the disparate forces from the justification-suppression model (justification vs. suppression; Crandall and Eshleman, 2003) that influence the display of discriminatory behaviors, there is potential to overlook the gravity of the problem of bias in interview ratings as trivial, when in fact meaningful differences may have just been obscured. For example, results from Levashina et al.’s (2014) meta-analysis on sub-group differences in structured interviews were reported in a manner that collapsed on the applicants’ capability or interview performance level – a consideration that may justify discrimination at low levels of performance and suppress it at higher levels. Had researchers taken this perspective, they may have uncovered more evidence of race-based differences at low vs. high levels of interview performance. This distinction is important in explaining how the expression of modern bias may be masked by traditional effect size comparisons of average group differences.

With this project I hope to first, shed light on how bias exists between subgroups on a range of performance (low to high) by drawing on the justification-suppression model (Crandall and Eshleman, 2003). I propose to do so by experimentally manipulating interview performance
level to see if this factor influences the degree to which race-based discrimination is expressed. Should a justification suppression model explain modern bias in such a manner, this overlooked area of research (Ruggs et al., 2013) might garner renewed interest from those eager to address lingering issues with race-based discrimination in the employment hiring process. Second, this project will evaluate the efficacy of two common diversity trainings (goal setting and perspective taking) in their ability to reduce bias in interview ratings. The diversity training literature has provided support for these two trainings as viable methods for diverse others to work together effectively (Bezrukova, Jehn, and Spell, 2012; Todd, Bodenhausen, Richeson, and Galinsky, 2011; Wexley and Baldwin, 1986); however, it is unclear which of the two trainings might be more efficacious (Bezrukova, Jehn, and Spell, 2012) for reducing discrimination in interview ratings. As I/O Psychologists, it is simply not enough to detect bias in organizational systems; we need to be a part of the solution also. A unified research effort tying together the detection of illegal or ethically poor employment decisions with possible remediation aligns to how organizations operate (Castilla, 2008; Gillespie and Dietz, 2009), and reflects the approach of the current project. Accordingly, each study has the capability to contribute to the literature uniquely and together by attempting to (a) identify and (b) rectify discrimination against Arab applicants in employment interviews.
Chapter 2

Literature Review

Stigmatized Individuals and the Various Expressions of Discrimination

Stigma is a term that represents a deeply discrediting attribute or individual characteristic (visible or not) that may be used as the basis for devaluation or discrimination (Goffman, 1963). Stigmas can encompass many different aspects of an individual and fall on a variety of dimensions. Examples of stigma that are often subject of research in the field of interview bias are race (Brief, Dietz, Cohen, Pugh, and Vaslow, 2000; Orpen, 1982; Riach and Rich, 2002), sex (Gilmore, Beehr, and Love 1986; Gallois, Callan, and Palmer, 1992), obesity (Agerström and Rooth, 2011; Bell and McLaughlin, 2006; Hebl and Kleck, 2002), physical unattractiveness (Dipboye and Colella, 2005; Hosoda, Stone-Romero, and Coats, 2003), both physical and developmental/intellectual disabilities (Hayes and Macan, 1997), and ethnicity (Kaiser and Pratt-Hyatt, 2009; Rakic, Steffens, and Mummendey, 2011). The defining attribute shared amongst these studies is that observation of the applicant’s stigma triggers emotional reactions in the interviewer that include uncertainty, discomfort, anxiety, and perceived danger (Bos, Pryor, Reeder, and Stutterheim, 2013; Major and O'Brien, 2005). These emotional responses are consistent with reactions to threat and fear (Blascovich, Spencer, Quinn, and Steele, 2001). This fear has an evolutionary origin as many physical stigmas would be perceived to be indicators of a disease or a threat (Park, Schaller, and Crandall, 2007; Ryan, Oaten, Stevenson, and Case, 2012). Stigmas will therefore initiate schemas (a mental structure of preconceived ideas) that invite behavioral, cognitive, and affective reactions in the interviewer including avoidance, fear, disgust, and the activation of stereotypes (Amodio and Devine, 2006; DiMaggio, 199; Neumann and Strack, 2000; Zajonc, 1980).
Recent reviews of the interview literature have highlighted various types of visible stigmas that can influence interviewers' judgments: pregnancy, visible religious marks, tattoos, obesity, gender, age, race, physical disability or a facial stigma (e.g., Dipboye and Johnson, 2013; Huffcutt, 2011; Macan, 2009; Macan and Merritt, 2011). Importantly, Pryor, Reeder, Yeadon, and Hesson-McInnis (2004) noted that visible stigmas can be controllable or uncontrollable, and that people's negative reactions towards stigmas were stronger when the target was perceived to be responsible for the stigma. Visible and uncontrollable stigmas that could play a role in the interview include gender, age, race or ethnicity, physical disability or a facial stigma (e.g., a port-wine stain). Stigmas that can be perceived as controllable include pregnancy, visible signs of religion or religious beliefs, tattoos, and being overweight or obese (although this can be uncontrolled, for instance when caused by illness, overweight individuals are often perceived as responsible for their condition; Randall et al. 2017). The activated reactions to stigmas can vary greatly depending on the attributable characteristics the interviewer holds – some may choose to express their cognitions or behaviors subtly or overtly in formal or interpersonal manners (Cuddy, Glick, and Beninger, 2011; Dovidio and Gaertner, 2004).

One visible and less controllable stigma that has received minor attention in the interview bias literature is the stigma attributable to the Arab ethnicity (King and Ahmad, 2010). When the selection bias research field was more popular a few decades ago, the overwhelming majority of research on intergroup biases defaulted to investigating Black-White differences (Crosby, Bromley, and Saxe 1980; Dipboye and Colella, 2005). Although many Americans identify with a Judeo-Christian heritage, a 2010 Census bureau report placed representation of Arabs at over 1.5 million individuals living in the United States, while current private industry estimates suggest this number had risen to three-four million people by 2018 (Gallup Research Center, 2018; Pew
Additionally, presenting with an Arab status is qualitatively different than presenting as Black or Hispanic. Arabs are perceived as “less American” and less acculturated than other ethnic groups (Suleiman, 2010). Another qualitative difference comes from the aftermath of Arab perceptions post 9/11, in which many Americans held and continue to hold Arabs in hostile and fearful regards due to being seen as a perceived threat. Further evidence of this can be found within the American legal system where the frequency of court cases and size of punitive measures against Arabs have markedly grown from before 9/11 to post-9/11 (Suleiman, 2010). Indeed, the opportunity to study these qualitative differences have gone largely overlooked in the interview bias literature. Despite the lack of focus on the Arab stigma in an interview context, the work of Fiske, Cuddy, and Glick (2007) found, through the lens of social cognition, that the Arab ethnicity is one of the most disparaged groups across social interactions—displaying stronger effects than people of color more historically represented in selection research (e.g., Black, Hispanic).

Social cognition focuses on how people process, store, and apply information about other people in different scenarios (Fiske, 2004). Some of these processes and applications have been found to apply across a vast array of social interactions, earning the term of universal social cognition dimensions (Aronson, Wilson, and Akert, 2010; Fiske, 2004). Warmth and competence are two such dimensions that have been found to account for 82% of the variance in perceptions of everyday social behaviors (Wojciszke, Bazinska, and Jaworski, 1998). Thus, these two dimensions serve as broad categories by which individuals tend to judge or evaluate other individuals and groups. The prevalence of these two dimensions is thought to be explained by the notion that people are evolutionarily predisposed to first assess someone’s intent to either do harm to them or help them (warmth) and second, to judge the person’s ability to act on that
perceived intention (competence). This can be considered an evolutionary explanation of mechanisms that emerged even in the time of early human history, when neighboring outgroups evaluated one another with the goal of self-preservation. In these kinds of group dynamics, there is utility in knowing the valence of outgroup members’ intent (helpful or harmful) as well as their capability to make good on such intentions.

Based on this research, Fiske and colleagues (2002) developed the Stereotype Content Model which suggests that all outgroup members can be classified with regard to warmth and competence. Consistent with this understanding, cluster analysis of ratings of 25 stereotyped target groups (including Arab) by participants from 18 nations and found support that the dimensions of warmth and competence account for variability in the content of stereotypes (Fiske, 2004; Fiske et al., 2002). For example, stereotypes of working women are characterized by perceptions of high competence and low warmth and stereotypes of mothers are characterized by low competence and high warmth (Cuddy, Fiske, and Glick, 2004). Further findings clustered the evaluations of warmth and confidence towards the Arab ethnicity to be on par with that of welfare recipients and the homeless, earning hostility from outgroup members (Fiske, Cuddy, and Glick, 2007). Out of all of the racial or ethnic subgroups examined in this body of work, Arabs were rated with the combined least warmth and competence across all samples. Furthermore, Asians were viewed as more warm and more competent than Arabs, earning a perception of cold competence – meaning they are seen as skillful but with suspect intentions. Blacks were rated as more warm than Arabs (competence ratings were either comparable or much higher depending on whether the target was “Poor Blacks” vs. “Black Professionals”). These poor warmth and competence ratings have real world consequences. Fiske, Cuddy, and Glick’s research as well as further investigations have also gone on to classify Arabs in the
category most likely to receive contempt and active harm (Fiske, Cuddy, and Glick, 2007; Park, Felix, and Lee, 2007). In organizations, management may overlook an Arab individual for promotion due to a lack of perceived competence; recruiters and interviewers might exclude an Arab candidate in the hiring process due to a perceived lack of warmth or sociability. Further still, these two negative attributions may impede an Arab individual’s performance in group work settings, through both interpersonal and task related hindrance (Cuddy, Glick, and Beninger, 2011). Beyond the newsworthy incidences of overt discrimination (Rolander, 2019; O’Malley, Barlass, and Begley, 2019; Rodriguez, 2019) Arabs are subject to daily acts in the American workplace that are hallmark of systematic disparate treatment (Derous, Ryan, and Nguyen, 2012; King and Ahmad, 2010).

Although incidences of overtly prejudiced behaviors toward many stigmatized targets are rarer in contemporary American society than they were a few decades ago, subtle indicators of differential treatment persist in the form of “everyday discrimination” (Deitch et al., 2003; Dovidio and Gaertner, 1986; McConahay, 1983). Researchers have used a variety of techniques to assess such subtle behaviors, including measures of interpersonal nonverbal behaviors such as social distance (Word, Zanna, and Cooper, 1974), facial expressions (Ruscher, 2001), eye contact (Dovidio et al., 1997; Jones and Cooper, 1971), and smiling (Biernat and Vescio, 2002). These behaviors are often aligned with unconscious attitudes or implicit associations that link targets of stigma with negative words, images, and outcomes (Greenwald and Banaji, 1995; Greenwald, McGhee, and Schwartz, 1998).

Only recently have researchers conceptualized discrimination found in organizational settings to be both formal and interpersonal in nature (Hebl, Foster, Mannix, and Dovidio, 2002). Formal discrimination involves behavior that is explicitly prohibited by laws or company
policies, and can include refusing to greet, help, hire, or train stigmatized customers or employees (King, Shapiro, Hebl, Singletary, and Turner, 2006). Interpersonal discrimination is understood as biased behavior that is not expressly illegal nor a job task. Interpersonal discrimination is verbal, nonverbal, or even paraverbal in nature. Examples of this include shortened interactions, decreased eye contact, decreased smiling, and rudeness (Hebl et al., 2002). In a series of experimental field studies, Hebl and colleagues found evidence of interpersonal discrimination toward targets who were seemingly gay and lesbian (Hebl et al., 2002), obese (King et al., 2006; Randall et al. 2017), or pregnant (Hebl et al., 2007). Additionally, King and Ahmad (2010) found that participants were more willing to discriminate against Muslim applicants for a job due to the perception of having less warmth than applicants of no religion. These workplace events are of special note in the discrimination literature due to the influential outcomes associated with these events such as applicant denials, terminations, lost wages, and loss of promotion advancement opportunities.

Modern Bias in Interviews

The modern understanding of discrimination involves both verbal and non-verbal behaviors that occur in social interactions with stigmatized persons (Dovidio and Gaertner, 2000; Hebl, et al., 2002). Stigmatized persons are classified as possessing characteristics that are seen identifiable and undesirable from others (Goffman, 1963). If interviewers react to and treat stigmatized individuals differently than non-stigmatized persons during the interview there are many poor resulting outcomes, including a potential miss on hiring top talent, costly legal challenges, having a poor social corporate image, as well as a loss in innovation within the company (Lambert, 2016; Levashina et al., 2014). Additionally, the differential treatment from interviewers could cause stigmatized groups to modify their job search behavior and
 interviewing behavior to their own detriment, thus creating and perpetuating a self-fulfilling prophetic type effect to disadvantaged subgroups (Hakim, Kurman, and Eshel, 2017; Taggar and Kuron, 2016). With these very real and costly consequences, it is surprising to find the science and practice not doing more to meet the challenge.

Many reviews of employment interview research have supported the idea that adding structure to the interview process can enhance the validity of interviewer judgements without the typical adverse impact tradeoff usually found for cognitive ability tests (Huffcutt and Culbertson, 2011; Huffcutt and Woehr, 1999; Posthuma, Morgeson, and Campion, 2002). Therefore, there is an expectation that standardization and structure to the interview process may help protect minority applicants from unfair evaluations (Strauss, 2018). Unfortunately, surveys show that despite the high overall frequency of interview usage, managers, human resource professionals, and organizations at large infrequently use structured interviews (Huffcutt and Culbertson, 2011; Lievens and De Paepe, 2004; Simola, Taggar, and Smith, 2007). Researchers and practitioners alike have had little success in promoting the use of structured interviews (e.g., Dipboye, 1997; Lievens and De Paepe, 2004). Instead, the trend appears to be the utilization of structured interview components, creating the broad term of semi-structured interviews. For example, an organization may ask the same questions to every candidate, but not base those questions off an established job analysis, thus utilizing only some, but not all aspects of a structured interview (Levashina et al., 2014; Lievens and De Paepe, 2004). Even in such settings, maintaining the standardization of the selection interview can be difficult in practice. When standardization becomes variant, it is possible for individual biases to interfere with the interview process (McKay and Davis, 2008).
The research on interviewee demographics and other characteristics on interviewer judgments is an established topic of consideration (Arvey and Campion, 1982; Harris, 1989; Macan, 2009; Posthuma, Morgeson, and Campion, 2002). Early studies mostly examined the direct effects of gender, race, or ethnicity on interview and hire-ability ratings. These early studies found small and inconsistent effects, leading some authors to suggest that stigmatized demographic factors may not substantially influence interviewers’ decisions (Macan, 2009; Posthuma, Morgeson, and Campion, 2002). Perhaps, this conclusion is one reason for the lack of attention in today’s journals. However, additional consideration is needed given more recent research. For example, Roth and colleagues (2002) found sizable Black-White differences for the use of the interview as an initial screening and Quillian et al.’s (2017) findings suggest hiring discrimination remains stable since 1989. Additionally, the limitations of previous investigations calls into question their interpretations and meaning. Effect size estimates were calculated using restricted samples (e.g., job incumbents or job applicants who had passed a previous selection test), resulting in standardized ethnic group differences not reflective or directly applicable to the entire job applicant population. Most surprising from these investigations is that mean group “differences were observed despite the fact that the interview was structured, with job-related questions based on job analyses, and trained interviewers” (Roth, Van Iddekinge, Huffcutt, et al., 2002, p. 375). This suggests that existing best practices to standardize the interview process in order to not disadvantage protected subgroups may not be enough to prevent race discrimination in employment interviews. Beyond standardization, the next facet to tackle involves the associated reasons why a minority receives differential treatment or scoring in an interview process. To that end, interviewer’s motivations to express or suppress their biases are explored.
Justified and Suppressed Biases

As mentioned previously, in Crandall and Eshleman's (2003) Justification-Suppression Model (JSM) of prejudice, people manage their biases by filtering the expression of prejudice through a process of suppression and justification. Specifically, when meeting a new person, one spontaneously recognizes if the person belongs to a stigmatized social category (e.g., the person is African American). This automatic categorization automatically activates stereotypes and prejudiced attitudes (e.g., this person is a threat; Devine 1989; Macrae, Bodenhausen, and Milne 1995). However, most people do not wish to express their prejudicial attitudes because their self-concept or social norms clash with these feelings, so they attempt to suppress activated stereotypes and prejudice (Crandall and Eshleman, 2003; Hebl and Dovidio, 2005). As a result, in-group members are not likely to knowingly commit overt acts of discrimination. This is especially the case when stigmatized individuals present stereotype-inconsistent cues that encourage suppression of the prejudicial attitudes. Such cues challenge stereotypes about one’s stigmatized subgroup; for example, an Arab man might challenge the stereotype that Arab people are threatening by smiling a lot or being overly courteous, encouraging the suppression of prejudicial attitudes in those he meets on the street or on the job. According to the JSM, then, individuals express prejudice only when there is sufficient justification - when they can express prejudice without feeling self-directed negative affect (guilt, or regret) or fearing social reprisals. In other words, the cues that warrant suppression are absent and are replaced instead by cues that trigger justification of discriminatory behavior. So if the emotional displays and behavior of the Arab man from the example above were instead furtive or standoffish (perhaps due to past negative experiences with strangers on the street), passersby may feel justified in thinking or acting in a prejudicial manner that reinforces the stereotype that Arabs are dangerous (e.g.,
scowling, crossing the street to avoid him). In fact, when expressions of prejudice are adequately justified, they may not even be recognized as prejudice (Crandall and Eshleman, 2003). Although Crandall and Eshleman originally focused on white American’s perceptions to establish the associated justifications and suppressions in the JSM, future studies were able to find evidence for tenets of the JSM within other stigmatized population’s perceptions as well (see Correll et al., 2007; Kahn and Davies, 2011).

King et al. (2006), over the course of three studies, exhibited this justification-suppression effect in an experimentally manipulated context with obese individuals. When customer service employees were exposed to obese shoppers in prejudice justification conditions (where obese actors promoted the view that their obesity was due solely to their own choices), they exhibited more interpersonal discrimination to the overweight shoppers than average-weight shoppers. However, in prejudice suppression conditions (where obese actors promoted the view that obesity is due to external factors such as genetics), employees engaged in less interpersonal discrimination. Similarly, Randall et al. (2017) found helping behavior towards obese individuals could be predicted by the moderating justification or suppression cues towards obesity perceptions. Simply put, individuals were more likely to suppress discrimination in order to help and be polite toward overweight individuals who challenged the stereotype that overweight people are lazy (by wearing a 10k finisher shirt), than when overweight confederates reinforced weight-based stereotypes (by eating a DQ Blizzard while asking for help). This justification suppression model has also found support in explaining attitudes and behaviors towards gay rights (justification of discrimination through religious beliefs; Hoffarth, Hodson, and Molnar, 2017) sexual assault on minority woman (justification through victim blaming; Katz, Merrilees, Hoxmeier, and Motisi, 2017), and Muslims (justification of discrimination through perceived
Muslim views of gender equality; Moss, 2017). In sum, across many stigmatized groups, discrimination is more likely to occur when a scapegoat or justification factor is provided, and less likely to occur when a stereotype-inconsistent cue or suppression factor is provided.

Consistent with Crandall and Eshleman’s (2003) theory, workplace research has shown that racial discrimination is most likely to occur in the selection process when it can be justified on non-prejudicial grounds. For instance, studies have shown that racial bias against African-American job applicants are more likely to occur when their qualifications are rather undistinguished, or ambiguous (Dovidio and Gaertner, 2000; Hodson, Dovidio, and Gaertner 2002; McConahay, 1986). Only then can the evaluator justify a negative evaluation on non-racial grounds (e.g., the applicant simply was not impressive). In contrast, racial bias is not likely to occur - that is, whites suppress bias against African Americans - when the African-American applicants are clearly well qualified (Dovidio and Gaertner, 2000; Hodson, Dovidio, aand Gaertner 2002; McConahay, 1986). Therefore, connecting this past research with the Stereotype Content Model (Fiske et al., 2002), it could be that the competence dimension that is used to stereotype outgroups may serve as a justification or suppression factor at different ends of the continuum. That is, an applicants’ high degree of competence or good performance in an interview may be seen as a suppression factor that prevents an evaluator from exhibiting bias in their interview ratings (i.e., the applicant is highly qualified, therefore suppressing the evaluator’s prejudicial views of low competence). Consequently, we would expect that at high levels of interview performance there would be less of a difference in interview ratings between minority and non-minority candidates. In other words, good performance suppresses discrimination. Conversely, if a stigmatized individual inadvertently reinforces stereotypes that his or her outgroup is low on the dimension of competence by performing at a lower level during
an employment interview, the evaluator may feel more justification for providing harsher ratings for the stigmatized applicant (i.e., the applicant is very unqualified, therefore justifying the evaluator’s prejudicial views of low competence). Thus, discrimination may be more likely at lower levels of interview performance for groups stereotyped as low in competence such as Arabs. Subsequently, we would expect to see more discrepant scores between minorities and non-minorities at this end of the competence continuum.

*Hypothesis 1:* Arab interviewees will be rated less favorably compared to White interviewees and the magnitude of this mean group difference will decrease as interview performance increases.

Figure 2 presents an illustrative display of the relationships hypothesized in Hypothesis 1 showing that the magnitude of mean score differences between White and Arab interviewees decreasing as interview performance increases. This research design is novel in regards to the examination of bias in interviews and affords the ability to look deeper into the structure of bias along the performance continuum. This in turn may contribute clarity to an otherwise muddied topic as well as offer practitioners rich data to incorporate into interviewer diversity training in order to more appropriately identify and address problems of bias in employment interviews.
Figure 2. Hypothetical illustration of directionally hypothesized relationships in one graph. High performance conditions are not necessarily expected to be equal, rather the differences will be less pronounced.

Diversity Training: A Possible Remedy for Bias in Employment Interviews

Due to the increase in discriminatory complaints in the workplace as well the prevalent biased attitudes to certain minority groups (e.g. increased contempt towards Arabs), finding new situations in which differential treatment is more likely to occur is important to understanding modern bias. In certain approaches to selection (e.g. multiple hurdle, top-down ordered lists), biased interview scores for stigmatized applicants may represent the difference between employment and unemployment. However, only finding the situations under which such discrimination is more likely is not enough. Similarly, for a doctor to merely diagnose a disease is not enough: treatment is required. Thus, research is needed to continue to develop and examine strategies that may reduce the frequency and magnitude of differential treatment in the workplace. Many of the suggestions of past research to reduce discrimination place the burden on the stigmatized group to display suppression cues, but such approaches are not wholly fair nor do they impact the underlying mechanisms. This investigation will go one step further by
evaluating methods to deter discriminatory ratings in interviews regardless of the presence or absence of suppression cues in the applicant.

The justification-suppression model of bias (Crandall and Eshleman, 2003) gives insight to when bias might and might not occur, and supplies some tactics for minority groups to reduce prejudicial treatment. King et al. (2006) and Randall et al. (2017) found that through targeting the justifications for prejudicial treatment, stigmatized individuals can reduce the severity of the disparate treatment they experience. That is to say, stigmatized individuals can challenge stereotypes associated with their stigmatized group in order to promote suppression of others’ would-be discrimination (Randall et al., 2017). For example, King and colleagues (2006) were able to manipulate how obese individuals were treated by utilizing cues such as a visual cue of a soft drink (high justification = ice cream drink; low justification = diet soda), indicating exercise (high justification = friends ran half marathon; low justification = target ran half marathon) and expressing whether they are on a diet or not (high justification = no diet; low justification = on a diet). Similarly, Randall et al. (2017) altered these conditions by using only visual cues (e.g. suppression cue: 10k race t-shirts, justification cue: ice cream drink) and never calling direct attention to the cues in the manipulation. One could imagine how interviewees could utilize the same strategies when they interview. An obese person in an interview might mention their healthy habits (exercise or diet) to the interviewer, someone with tattoos might cover them up, a minority of a foreign background might make an effort to wear traditionally American business clothes. Or, as could be argued if the results of this study support Hypothesis 1, Arabs cannot be just equally competent to non-Arab applicants in order to be equitably evaluated in an employment interview; instead, they would have to outperform the competition in order to get a fair shot. All of these are considered efforts exerted on the target’s behalf to reduce the
justification an interviewer might feel to discriminate against them. Given such recommendations, however, researchers and practitioners may make the mistake of placing the burden of responsibility to reduce prejudice on the discriminated rather than the discriminators (King et al., 2006; Randall et al., 2017). By contrast, diversity training may provide a more suitable route for suppressing prejudice as the burden to reduce the prejudicial treatment is placed back upon the discriminator, or in this case, the interviewer.

Diversity training has been referred to as an effort to facilitate positive intergroup interaction, reduce prejudice and discrimination, and generally teach dissimilar others how to work together effectively (Bezrukova, Jehn, and Spell, 2012). Diversity training is distinct from most other traditional workplace training in that it deals with subjective and emotionally-laden topics, such as one’s presumably ingrained attitudes about stigmatized groups (Hanover and Cellar, 1998). Whereas most training programs focus on the skill development and cognitive outcomes of training, diversity trainers tend to focus on attitudes and emotions as key training outcomes because attitudes toward specific outgroups and diversity at large are likely developed prior to training. As a result, diversity training sessions tend to be more emotionally and politically charged as a result (Paluck, 2006). This polarizing effect may in part be responsible for the observed backlash where individuals sometimes react negatively against diversity training, producing the opposite of desired training effects (see Legault, Gutsell, and Inzlicht, 2011). Diversity trainers hope to avoid this pitfall, and instead promote organizational compliance, social harmony, and inclusion (Rossett and Bickham, 1994). Meeting these goals is predicated on the belief that not only will the stigmatized and non-stigmatized individuals benefit from training, but that organizations also benefit by improved work attitudes, culture, and climate, increased creativity and innovation, and reduced negative consequences of
discrimination such as employee turnover and lawsuits (Bezrukova, Jehn, and Spell, 2012; Simons, 1992). Thus, the most important objective of diversity training is for groups of diverse individuals to learn how to work together effectively, which should lead to benefits for stigmatized and non-stigmatized individuals, as well as organizations as a whole (Bezrukova, Jehn, and Spell, 2012).

The generalized operationalization of diversity training programs has allowed for various implementations to occur. A recent meta-analysis (Nittrouer, Hebl, and Oswald, 2016) revealed that there was no strongly preferred form of course instruction in terms of training design – rather the mixed preferences involved lecture (21%), reading (19%), discussion (19%), videos (10%), and role plays (10%). Bezrukova and colleagues’ (2012) review found that 91% of diversity training studies either did not specify their delivery methods or utilized multiple methods. Although this should not be surprising given that most training research concludes that multiple methods are more effective than using a single method (e.g., Arthur, Bennett, Edens, and Bell 2003; Tannenbaum and Yukl, 1992), using multiple methods per study has hindered research on isolating the effectiveness of each activity. In a qualitative review of the effects of diversity training, Kulik and Roberson (2008) found it to be effective in terms of improving overall attitudes toward diversity; however, the effects were inconsistent when considering attitudes towards specific stigmatized groups. The authors go on to suggest that trainees tend to perceive themselves as having more dexterity in interacting with diverse populations after training has taken place – although, there is little evidence to support the idea that diversity training actually improves attitudes and, more importantly, behaviors toward stigmatized groups (Kulik and Roberson, 2008). In the largest and most recent meta-analytic review on the effects of diversity training, Bezrukova, Spell, Perry, and Jen (2016) found that there is a strong effect of
diversity training on cognitive learning and reactions to training (Hedges g of 0.38) but more modest effects for attitudinal and behavioral learning. Interestingly, many of the diversity training programs fell short in demonstrating effectiveness in their analysis, however the authors were able to establish that successful diversity training does occur. The positive effects of diversity training were greater when training was complemented by other diversity initiatives, targeted to both awareness and skills development, and conducted over a significant period of time (Bezrukova et al., 2016).

Given that diversity training initiatives work to improve diversity attitudes and behaviors so that dissimilar individuals work inclusively with one another (Bezrukova, Jehn, and Spell, 2012), this lack of evidence in the literature for attitude and behavior change (Bezrukova et al., 2016; Nittrouer, Hebl, and Oswald, 2016) is distressing for diversity training science and practice. The diversity training literature typically measures intentions or attitudes as outcomes (Combs and Luthans, 2007; Kraiger, Ford, and Salas, 1993), stopping short of assessing its effects on actual discriminatory behavior. This study plans to address this shortcoming by directly measuring whether people will discriminate or not on interview ratings following a diversity training, thus gauging the direct rather than inferred efficacy of the trainings. Applying this focus addresses a key limitation in the diversity training space (Bezrukova, Jehn, and Spell, 2012; Kraiger, Ford, and Salas, 1993; Kulik and Roberson, 2008; Nittrouer, Hebl, and Oswald, 2016).

This section will attempt to summarize the efficacy of two different methods that the literature has shown to have the most positive impact on diversity related outcomes. Each method raised from the literature—goal setting and perspective taking—are both applicable to
targeting attitudes toward specific stigmatized groups, yet each is based on a different theoretical underpinning.

**Perspective Taking**

Perspective taking in diversity training can be defined as actively considering the psychological experiences of others, with an emphasis on how those experiences may differ from our own (Todd et al. 2011). This is most frequently accomplished by utilizing some sort of manipulation (video, role-play, discussion, recording) that elicits empathy for the target from the participant (Roberge and Dick, 2010). Empathy is conceptualized as a process through which people identify with and understand another's situation, feelings, and motives. As recently pointed out by Milton and Westphal (2005), empathy is critical in explaining the process by which people learn about each other, incorporate others' identities into their own self and, as a result, arrive at a better understanding, tolerating and cooperating with one another. Batson and colleagues (1997) demonstrated how this plays out through a perspective taking paradigm. In several studies, participants listened to scripted interviews from women with AIDS and homeless men. Those participants that were prompted with statements like “try to feel what the interviewee on the tape was feeling” reported more positive attitudes towards the respective groups than participants who were instructed to be objective.

Within a social identity theory (Tajfel and Turner, 1986) framework, perspective taking is thought to be effective because it breaks down psychological barriers between in-groups and out-groups. Indeed, individuals tend to categorize themselves and others into in-groups and out-groups based on meaningful social variables such as ethnicity and sexual orientation (Hogg and Terry, 2000). Although this categorization process has the potential to improve in-group self-esteem via social comparison (Tajfel and Turner, 1986), it also has the potential to produce in-
group favoritism and bias toward out-groups that could result in prejudice and ostracism (Dudley and Mulvey, 2009). Thus, perspective taking reduces prejudice by requiring individuals to think about what it would be like to be a member of a different group, which serves to break down in-group versus outgroup barriers and more generally reduces an “us versus them” mentality (Galinsky, Ku, and Wang, 2005). Research indicates that taking the perspective of another person can lead trainees to view others more positively. For example, one study found that when people took the perspective of an individual from a stigmatized group, they tended to express more positive attitudes toward that group overall (Batson et al., 1997). Another study replicated this finding and also found that taking the perspective of a stigmatized individual decreased the extent to which participants used stereotypes to judge members of that group (Galinsky and Moskowitz, 2000). Galinsky and Moskowitz (2000) go on to suggest that perspective taking works through reducing the activation of discriminatory thoughts and scripts, and replacing them with empathetic cognitions and feelings. In this way, we can think of perspective taking as reducing the thoughts that would occur in justifying bias through Crandall and Eshleman’s (2003) Justification-Suppression Model. That is to say, perspective taking works because it replaces the cognitions associated with justifying one’s bias with thoughts and feelings of understanding about the target group’s experience. Imagine being forced to use a wheelchair for a day to navigate a college campus that was built with limited accessibility. Following that experience, the next time you see someone wheelchair bound navigating to class, you will remember your own experiences and may be more likely to assist that person.

Goal Setting

Another method practitioners have used in diversity training involves participants setting diversity-related goals that encompass a specific stigmatized group. Goal setting theory states
that performance will be improved to the extent that participants set specific and challenging, but achievable goals (Locke, Shaw, Saari, and Latham, 1981). In diversity training, goal setting theory can be incorporated by asking trainees to set high quality goals aimed at promoting diversity within an organization. Goals are thought to have their positive effect by driving motivation and guiding behavior (Locke and Latham 1990). Importantly, goals are also thought to be an effective transfer of training strategy that can enhance or at least retain the effectiveness of a training session long after it has concluded (Latham and Seijts, 1997). This is of particular importance in the realm of diversity training, where training sessions are generally infrequent and may be quite short.

Self-set goals are a particularly useful strategy for diversity training because they can provide trainees with concrete objectives and solidify the behaviors and attitudes necessary to meet the self-set objectives (Madera, King, and Hebl, 2013). Some examples of goals related to diversity might include refraining from using derogatory words when talking about or to minorities, attending minority alliance meetings, or reading literature about modern minority issues. Over time, as individuals set increasingly difficult goals, they will be exposed to alternative evidence about the minority groups. Such goals related to diversity might lead individuals to engage in fewer prejudicial behaviors, more supportive behaviors, and to create new thought and attitudinal patterns toward the target minority groups (Madera, King, and Hebl, 2013). The integration of how goal setting might help reduce the expression of prejudice through the lens of the Justification-Suppression Model is rather intuitive. A key underpinning of the JSM as put forth by Crandall and Eshleman (2003) posits that in contemporary society, people tend not to identify with prejudicial and discriminatory self-concepts, nor do they wish to be seen as such by others. In a goal setting framework, this training utilizes that motivational energy and
scaffolds it with tools (e.g., self-regulatory strategies to monitor goal progress; Sitzmann and Ely, 2011) to achieve legitimate behavior and attitude change (Locke and Latham, 2002). Depending how the goal is chosen, this training strategy can either reduce the activation of justification factors (e.g., reading literature to understand why a certain minority group dresses a certain way, thus removing that factor in the future) or increase the activation of suppression factors (e.g., becoming more self-aware of certain discriminatory cognitions in key interactions and choosing to remove them from awareness).

Empirical work has demonstrated the effectiveness of goal setting in diversity training and more general training contexts. Specifically, goal setting has been shown to produce positive changes in leadership and interpersonal skills training (Wexley and Nemeroff, 1975) and time management skills training (Wexley and Baldwin, 1986). Additionally, goal setting three months after an initial training has been shown to enhance the effectiveness of safety training nine months after the initial training had taken place (Reber and Wallin, 1984), which indicates the potential of this strategy to produce long-lasting attitude and behavior change. However, Richman-Hirsch (2001) found that goal setting leads to training transfer significantly more frequently when the transfer climate is supportive compared to when it is not. That is to say, if an organization wanted a diversity training to succeed then they would have to provide opportunities and support after training has ended for individuals to practice and apply the skills they are intended to learn (Tracey, Tannenbaum, and Kavanagh, 1995).

The use of goal setting to promote diversity is supported by a recent study (Madera, King, and Hebl, 2013) in which students were asked to set goals promoting acceptance of LGB individuals in addition to receiving standard diversity training while other students only received the standard training (i.e., a no-goal setting control). Results after three months showed that
students who had participated in the goal setting training exhibited more behaviors promoting the acceptance of LGB individuals than those who set no goals. After eight months, trainees who set goals during diversity training also reported more positive attitudes toward LGB individuals (Madera, King, and Hebl, 2013).

The literature’s findings regarding diversity training are promising but unclear regarding which approach would best address the problem of bias in interviews. Goal setting has the capacity to be more behavior oriented than perspective taking and so, goal setting may be more effective at reducing bias expressed in interview scores than perspective taking. Additionally, goal setting in general training environments has a well-established, large effect on learning and training transfer in general (Arthur et al., 2003; Beier and Kanfer, 2010; Sitzmann and Ely, 2011). However, diversity trainings overall should improve diversity attitudes and behaviors so that dissimilar individuals work inclusively with one another (Bezrukova et al., 2012). To that end, this investigation will test the efficacy of these two forms of diversity training to reduce bias across all justification or suppression conditions. If the diversity trainings are having the intended effect, individuals should rate the Arab applicants more similarly to their White counterparts, regardless of condition. Additionally, diversity training should reduce the bias that emerges more prominently when justification is offered for discrimination (i.e., at lower levels of interview performance, see Hypothesis 1). However, I predict that the goal-setting approach to diversity training will be more effective overall at reducing this discrimination relative to the perspective taking approach due to behavior-oriented modality and established history of goal-setting’s benefits for training. Further, within a JSM framework, goal setting might outperform perspective taking as a diversity training approach because setting goals to specifically reduce expressed bias even when individuals might feel justified to do so is a more direct intervention in
regards to the outcomes of interest in these studies (e.g. interview ratings). Perspective taking in contrast, relies on an individual’s elicited empathy to replace the biased thoughts and feelings, which in turn reduces the observed bias in interview ratings, if successful. Finally, perspective taking assumes that empathy and bias towards a stigmatized target are at least somewhat mutually exclusive which might not always be the case.

To reiterate, there is a lack of evidence in the literature to support the efficacy of diversity training to reduce discriminatory behavior. Instead, the literature typically measures intentions or attitudes as outcomes (Combs and Luthans, 2007; Kraiger, Ford, and Salas, 1993). To address this, I will directly measure whether people will discriminate or not on interview ratings following a diversity training, thus gauging the direct rather than inferred efficacy of the trainings. Applying this focus addresses a key limitation in the diversity training space (Bezrukova, Jehn, and Spell, 2012; Kraiger, Ford, and Salas, 1993; Kulik and Roberson, 2008; Nittrouer, Hebl, and Oswald, 2016). By directly manipulating the actor’s interview responses to have varying degrees of justification factors (e.g., low interview performance provides justification to discriminate, vs. high interview performance provides very little or no reason to discriminate), this study will further contribute by examining the impact of training in contextually disparate situations.

_Hypothesis 2a:_ People who participate in a diversity training will display less observed bias in interview ratings for Arab relative to White applicants compared to individuals who do not attend a diversity training across all levels of interview performance (low, medium, high).

_Hypothesis 2b:_ People who participate in a goal-setting diversity training will display less observed bias in interview ratings for Arab relative to White applicants compared to
individuals who participated in a perspective-taking diversity training across all levels of interview performance (low, medium, high).

*Hypothesis 3a:* People who participate in a diversity training will suppress their rater bias against Arab interviewees compared to people who participate in the control group even when justification factors are present (i.e., at low levels of competence).

*Hypothesis 3b:* People who participate in a goal-setting diversity training will suppress their bias against Arab interviewees significantly more than people who participate in a perspective-taking training even when justification factors are present (i.e., at low levels of competence).

**Current Study**

This investigation examined the expression of modern bias through the lens of the justification-suppression model (Crandall and Eshleman, 2003) in order to detect and explain possible discrimination Arabs may experience during employment interviews. This was done in two studies using participants recruited from Amazon’s Mechanical Turk. Both studies utilized the same materials for pre-recorded interviews of a male answering questions at low, moderate, or high levels of interview performance/competence—a variable that was manipulated between-subjects. The male’s ethnicity (White or Arab) was also manipulated between-subjects in both studies by using a resume that included a traditional Arab or White American sounding name along with education and work experiences in either Saudi Arabia or America, holding the actor constant. Ratings of interview performance were aggregated and evaluated for potential differences between the White and Arab conditions (Hypothesis 1). Study two attempted to evaluate strategies that might ameliorate observed differences in Study one. Additionally, Study two empirically tested the impact of the two diversity trainings on justified and suppressed bias.
In Study two, participants first completed an online diversity training (goal setting, perspective taking, or control – internet safety). Next, they had two weeks to practice and display the behaviors they learned in their training sessions. Finally, those participants were invited back at the conclusion of the two-week window to undergo the same procedure as in Study one, rating Arab and White applicants’ interview responses using videos with the same three levels of interview competence manipulated between-subjects. The results from this study were used to re-test Hypothesis 1 and to evaluate Hypotheses 2-3, as well.

Without accounting for the disparate forces from the justification-suppression model (justification vs. suppression; Crandall and Eshleman, 2003) that influence the display of discriminatory behaviors, there is potential to overlook the gravity of the problem of bias in interview ratings as trivial, when in fact meaningful differences may have just been obscured. The detection of the hypothesized biases will be replicated across two studies because the control group for Study two will be methodologically similar to Study one. Additionally, due to the strong negatively biased warmth and competence views towards Arabs found by Fiske et al. (2007), the strong post 9/11 discrimination statistics on Arabs (Fiske, Cuddy, and Glick, 2007; Gallup Research Center, 2017; Pew Research Center, 2018; Zogby International, 2007), as well as the relative novelty of Arabs as a researched sub-group (King and Ahmad, 2010), this investigation contributes to a much needed question in the literature concerning the extent to which bias toward Arabs may influence employment interview ratings.
Chapter 3

Method and Analyses

Study one Method

Participants

The participant sample for Study one was recruited from Mechanical Turk with the restriction of 21+ in age and part-time (20+ hours) to full-time employees who have reported conducting at least one interview in a work setting previously, as suggested by previous researchers on this topic (Buhrmester, Kwang, and Gosling, 2011; Paolacci, and Chandler, 2014; Rouse, 2015). A working 21+ age group sample was sought after because this sample of workers are more likely to have conducted and participated in an interview in a high stakes setting, thus yielding a sample that more closely approximates the population of interest. Additionally, these restrictions have yielded better data quality in terms of careless responding in previous studies (Paolacci, and Chandler, 2014; Rouse, 2015). Only those with IP addresses in the US were allowed to participate due to differential cultural views towards Whites and Arabs potentially confounding results. Payment to participants was $1 due to the entire procedure estimate for Study one taking approximately 20 minutes and no special skills were required. A priori power analyses set the sampling goal to 223 or more. This number was derived from inputting the 2x3 fixed effects ANOVA (special effects, and interactions) design into a power analysis (interaction effect size $f = .05$, power = .80, groups = 6, number of predictors = 3 [race, interview condition, and interaction effect], response variables = 1). Hebl et al. (2002) in a similar study found effect size estimates ranging from .08 to .19, so .05 was chosen as a conservative estimate from these findings. In order to achieve this result, an initial target sample of 418 was collected, examined for data quality, and then the study concluded due to achieving the minimum sample size.
requirements (Zhou and Fishback, 2016). Further data cleaning resulted in the removal of 99 total cases, as a result of failing attention checks (e.g. select “red” if you can read this; 18 people), failing manipulation checks (48 people), time on page (two seconds on page or less were removed; 29 people), and CAPTCHA verification (four people). This left 319 participants retained for analysis which surpasses the power analysis goal of 223. In this sample, 67.1% were male, 54.5% identified as Caucasian, 21.9% as African American, 8.2% as Native American, 7.2% as Hispanic or Latino, 4.7% as two or more races, and 3.5% chose to not self-identify. 84.3% of the sample work full-time with 15.7% working part-time. This sample averaged 7.4 (SD = 5.0) interviews conducted in the workplace setting prior to their participation in Study one.

Procedure

Participants who met the aforementioned criteria were presented with an informed consent form. Those who agreed to the terms of the study answered a questionnaire that collects their demographic data such as race, ethnicity, and age. Participants were then given a scenario to read in which they are an interviewer evaluating a candidate for a sales role and they will receive a brief frame-of-reference training to familiarize themselves with the interview rating scale and procedure (See Appendices A-C for materials). Next, participants reviewed a resume that had either a traditionally Arab or White American name, and their educational and work experience in America (White condition) or in Saudi Arabia (Arab condition). Despite these differences, the degrees earned, work tenure, roles and responsibilities were the same between the two resumes. Next, participants were presented with one of three interview conditions—low, medium, or high competence responses—presented via video. The same actor for both the Arab and White condition responded to three broad, behaviorally based interview questions (See
Appendix C for Arab manipulation materials). The actor’s observable ethnicity was manipulated through resume reviews altering names, educational experience, and work experience locations (see Appendix C). Using the same actor and videos for White and Arab conditions gives the investigation more experimental control and reduces the number of possible alternative explanations for score differences (e.g., due to differences in interviewees’ communication or physical attractiveness). Between each interview question the participant received a behaviorally anchored rating scale (BARS) to rate the candidate’s response. Interview questions and responses, along with the rating scale, are presented in Appendix A. In order to exercise experimental control over the justification and suppression factors, the scripts and recorded interview responses were first pilot tested and rated by 11 Industrial/Organizational psychology student and faculty SMEs to ensure that the responses for the targeted low, moderate, and high levels of performance meaningfully differed in the expected directions. Intra-class correlation coefficients (2,1) were calculated among the SME judges to assess the reliability of ratings based on the interview scripts and BARS for each question. A single measures ICC (2,1) of .58 suggested moderate reliability among judges (Koo and Li, 2016). Upon further investigation, there seemed to be some disagreement among raters between high and average responses to the interview questions, so I added more anchors to the respective BARS and made reference to them in the correlating responses to better differentiate medium vs. high performers. In this way, we have more confidence that observed differences in interview scores between White and Arab conditions are due to biasing factors rather than true differences in interview performance.
Measures

**Interview Ratings.** In order to create vignettes of interview performance at low, medium, and high levels, three generic interview questions were asked of the videotaped confederate and averaged to give to an overall interview score.

1. Describe a situation where you found yourself working on a project with a new co-worker. How would you make sure the project was done well?

2. Suppose you had many projects with rigid deadlines, but your manager kept requesting more paperwork, which you felt was unnecessary. This extra work was going to cause you to miss your deadlines. What would you do?

3. Suppose you were going to run an important business meeting with your manager and leadership because they want to learn more about your recent major successes in your sales job. How would you prepare for the meeting?

These questions were adapted from a scale meant to cover a domain of global professionalism (Hofmeister, Lockyer, and Crutcher, 2009). Participants were provided with a job description of a generic sales role for more context. Participants in the study saw three videos of a man in professional dress (the same actor was used for the White and Arab conditions) displaying answers that were behaviorally consistent across the three questions – consistent with a BARS rating of one, three, or five for each question. A composite score was then derived from these three ratings by averaging responses within-person. Cronbach’s Alpha was calculated across the three performance questions because even though each question is tapping into slightly different generic sales competencies, each participant should be viewing similarly low, average, or high answers. A Cronbach’s Alpha of .52 suggested sub-optimal internal consistency between these measures. Further investigation showed that correlations between question one
and question three (inter-item correlation = .20) were driving down the relationship, whereas question one and question two had an inter-item correlation of .36 and question 2 and question 3 had an inter-item correlation of .23. These low reliability estimates are not necessarily a concern due to each interview question tapping into related, but somewhat different dimensions of performance, and to the small overall number of items which can influence the alpha coefficient. Past research has found similar reliability estimates when assessing performance (e.g. .54 - .63, Catano, Darr, and Campbell, 2007; Motowildo and Dunnette, 1990) and so it was deemed that there is sufficient reliability to aggregate interview ratings and use this measure as a dependent variable in the study.

**Warmth and Competence.** To further examine the functioning of the manipulation, nine questions were asked about the applicant’s warmth and competence – five assessing competence and four assessing warmth. Taken from Fiske et al. (2002), an example of a competence question would be “As viewed by society, how confident are members of this group?” (1 = not at all to 5 = extremely). Similarly, for warmth “As viewed by society, how good natured are members of this group?”. Reliability estimates for these measures were adequate, although a bit lower for warmth compared to competence: Warmth α =.64, Competence α =.80.

**Political Orientation.** A single item, general political orientation question adapted from Everett (2013) was asked to participants with the phrasing “How liberal vs. conservative would you characterize yourself?”. Responses ranged on a 7-point scale from Far Right Conservative (1) to Far Left Liberal (7).

**Random Responding.** In order to guard against random responding and poor data quality, a random responding question was asked after each video response was played. An example of
such a question was “If you can read this, select strongly disagree”. Eighteen participants failed one or more of these checks and were excluded from analyses.

**Manipulation Checks.** Following the resume manipulations that showed the applicant to be of either an Arab background or White American background by changing the name, former job experience and education, manipulation checks were presented that asked “Where is (John/Mohamad)'s B.S. in Business Management from?” and “Where is (John/Mohamad)'s most recent work experience from?” with drop-down options containing the correct answer and a distractor response. Forty-eight people did not correctly identify the manipulation and were therefore removed from analysis.

**Analysis Strategy**

The software utilized for all analyses was IBM SPSS 26. I evaluated Hypothesis one by running a 2x3 fixed effects design Analysis of Variance (ANOVA) on interview ratings and examining the interaction effect of interview performance (three levels: low, medium, high) by ethnicity (two levels: Arab, White). A main effect of interview performance would confirm that my manipulation of the low, medium, and high levels of performance were successful in eliciting a biased response in interview scores, and a main effect of ethnicity would reveal any overall bias between the White and Arab manipulation. A significant interaction effect would indicate that the presence of ethnic differences differentially impacts certain levels of interview performance more than others, beyond the main effect of race or the interview competency manipulation. If a significant interaction was present, then simple main effects and the resulting effect size estimates (eta squared) comparing Arab and White interviewees at each of the three levels of performance would help illustrate the presence or absence of a justification-suppression effect with larger between-group effect sizes expected at the lowest level of competence.
compared to the other levels (Hypothesis 1). ANOVA normality assumptions were tested through the Shapiro-Wilk test and the homogeneity of variance assumptions were tested through Levene’s test; no violations were found.

**Study one Results**

Descriptive statistics and bivariate correlations of all study variables are presented in Table 1. These reveal that interview scores, warmth ratings, and competence ratings all share significant relationships with one another such that higher interview scores were associated with higher warmth and higher competence ratings, and vice versa. Additionally, political orientation did not share significant variation with interview scores but did share significantly negative relationships with warmth and competence ratings such that those that identified more conservatively rated applicants as more warm and more competent and those that identified more liberally rated applicants as less warm and less competent. Correlations between both forms of experimental manipulations and interview scores were non-significant.

Table 1

*Correlations Among and Descriptive Statistics for Key Study 1 Variables*

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interview Scores</td>
<td>3.14</td>
<td>0.85</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Warmth</td>
<td>3.50</td>
<td>1.00</td>
<td>.16**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>3.60</td>
<td>0.91</td>
<td>.31**</td>
<td>.37**</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Political Orientation</td>
<td>4.04</td>
<td>1.87</td>
<td>-.06</td>
<td>-.16**</td>
<td>-.20**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5. White/Arab Condition</td>
<td>1.50</td>
<td>0.50</td>
<td>-.07</td>
<td>.10</td>
<td>-.01</td>
<td>-.02</td>
<td>-</td>
</tr>
</tbody>
</table>

*Notes. N=319. For White/Arab condition, 1 = White condition, 2 = Arab Condition – correlations with this variable are point-biserial. Political Orientation coding – 1= Far Right Conservative, 7= Far Left Liberal. * p < .05. ** p < .01.*
To test Hypothesis 1, the 2x3 ANOVA on interview scores described above showed a significant main effect for the interview competence level manipulation, $F(2, 313) = 6.50, p < .01, \eta_p^2 = .04$, suggesting that raters were successfully able to distinguishing between performance conditions (low, moderate, high). However, the main effect of ethnicity $F(1, 313) = 1.12, p = .21, \eta_p^2 = .01$, and subsequent interaction effect $F(2, 313) = 0.97, p = .38, \eta_p^2 = .01$, were both not significant, suggesting that interview ratings did not significantly differ by ethnicity, nor did ethnicity and interview performance group interact to influence interview ratings. This suggests that participants rated the Arab and White applicant similarly across all levels of interview performance (low, moderate, high), failing to support the idea that Arabs would be rated lower than Whites and that this difference would be most pronounced at low levels of interview performance. Hypothesis 1 was not supported. Rather than averaging the three items for a composite interview score, additional item level tests of Hypothesis 1 (i.e., running the 2x3 ANOVA with each interview score as its own dependent variable) produced similar results in that the main effect of interview condition did produce significant results but neither the main effect of ethnicity nor the interaction effects of ethnicity and interview condition were significant. See Appendix D for more detail on item level testing.
Warmth and Competence data were also collected in Study one to examine not only outcome oriented results (interview ratings), but also to capture differences in how participants viewed Arab and White applicants. To further explore the nature of participant perceptions, interview ratings were replaced in the 2x3 ANOVA model first with competence perceptions as the dependent variable, and then again with warmth perceptions as the dependent variable. In the competence model, results surprisingly did not show a significant main effect for the interview performance level, $F(2, 313) = 0.29, p = .75, \eta^2_p < .01$, suggesting that raters were not judging the competence of the candidate substantially differently despite the differences in interview performance (low, moderate, high). Additionally, the main effect of ethnicity $F(1, 313) = 0.02, p = .88, \eta^2_p < .001$, and subsequent interaction effect $F(2, 313) = 2.33, p = .10, \eta^2_p = .02$, were both...
not significant, suggesting that competence ratings did not significantly differ by ethnicity, nor did ethnicity and interview performance group interact to influence competence ratings. In the warmth model, results did not show a significant main effect for the interview performance level, $F(2, 313) = 0.58, p = .56, \eta_p^2 < .01$, suggesting that raters were not judging the warmth of the candidate differently despite the differences in interview performance (low, moderate, high). Additionally, the main effect of ethnicity $F(1, 313) = 0.02, p = .90, \eta_p^2 < .001$, and subsequent interaction effect $F(2, 313) = 2.99, p = .052, \eta_p^2 = .02$, were both not significant, suggesting that warmth ratings did not significantly differ by ethnicity, nor did ethnicity and interview performance group interact to influence warmth ratings. See Appendix D for means and standard errors.

**Study one Discussion**

This investigation examined the expression of modern bias through the lens of the justification-suppression model (Crandall and Eshleman, 2003) in order to detect and explain possible discrimination Arabs may experience during employment interviews. Study one did not find support for the proposed justification of discrimination against the Arab, as opposed to White, candidate in the low-competence interview condition. Instead, suppression of discrimination or, more accurately a lack of discrimination, was evident at all levels of interview performance when comparing the Arab and White conditions. This may be due to a multitude of factors; the ethnic manipulation changing resumes may have been too weak of a stimulus to elicit the anticipated discrimination toward the Arab interviewee. Evidence of this can be found in the ANOVA on interview scores where the main effect of ethnicity is small and statistically non-meaningful ($\eta_p^2 < .001$). This study utilized a resume based manipulation while keeping the visual aspects of the interviewees identical in order to have greater experimental control over the
conclusions drawn from the results. This gain in internal validity did come at the cost of the strength in construct validity. Arab applicants very often present with a multitude of intersectional stigmatizing features that were not present in this study, such as cultural clothing, adherence to religious observations (e.g., hijab), accents, and even gender (King and Ahmad, 2010) all of which might produce a stronger link between the feature and someone’s Arab status. Alternatively, the manipulation may have instead (or also) caused demand characteristics that influenced respondents’ ratings. For instance, the negative bivariate relationship between political orientation, warmth, and competence ratings in Study one is counter-intuitive to previous work (Fiske and Durante, 2014), so more conservative participants may have been distorting ratings about the Arab applicant to be more favorable in order to avoid appearing overtly discriminatory as well as liberal participants rating the Arab applicant less favorably than those that that saw the white ethnic condition. In support of this, when the negative correlation between political orientation and warmth and competence scores is split by the ethnicity condition people saw, the white ethnic condition showed the relationship between political orientation and warmth to be $r(158) = -.18, p = .14$ and with competence to be $r(158) = -.22, p < .01$. While the Arab ethnic condition exhibited the relationship between political orientation and warmth to be $r(157) = -.20, p < .05$ and with competence to be $r(157) = -.26, p < .01$. The pattern of these results suggest political orientation shared a stronger relationship with warmth and competence in the Arab condition rather than the white condition.

Further evidence pointing to the weakness of the manipulation is found when examining warmth and competence ratings between conditions. Whereas previous studies have been able to establish the relatively low perceptions of warmth and competence for Arab individuals (Fiske et al., 2002; King and Ahmad, 2010), this study found no significant deviations in warmth and
competence between Arab and White conditions, nor indeed between high, moderate, and low interview competence conditions. There are several possible explanations for the observed results. Forty-eight people failed to correctly indicate where John/Mohamad’s education or work experience was from, which may be a signal that the ethnic manipulation was too weak to trigger stereotypical judgments. Participants may have been keyed into certain research questions of the study and choose to respond in socially desirable ways. Although social desirability was not directly measured, the association between political conservatism and higher scores on warmth and competence run counter to previous work in this area (Koch et al., 2016) Future studies should look to optimize their experimental designs towards stronger construct valid paradigms as well as explore novel paradigms around the expression of justified or suppressed bias. Despite the lack of supported findings of Study one, I wanted to re-examine these effects on a unique sample as well as move forward with the general aims of the investigation which also include exploring diversity training attempts to reduce potential discrimination due to justification pressures at low levels of interview competence. As these studies were pre-registered with a priori hypotheses, I chose not to alter my ethnicity manipulation for Study two as that would have invalidated the registration.

Study two Method

Participants

The participant sample for Study two were recruited from Mechanical Turk with the same restrictions applied in Study one, consistent with prior research recommendations. This was a two-part study separated by a two-week time period and participants were paid for their participation in each part of the study ($0.50 for time one; $1.00 for time two - $1.50 total), with payment pro-rated for those who did not return for Time two. Participants were contacted via
their registered email address on MTurk to participate in Time two of the study following two weeks after completion of Time one. To estimate the needed sample for a three-way interaction in ANOVA (race x interview competence x diversity training group), I estimated the effect size at $f = 0.05$, which was derived by the same methodology from Study one of conservatively estimating effects from prior research on similar variables (Hebl et al., 2002). The final sample size required was at least 295 participants. This number was arrived from inputting the 2x3x3 fixed effects (special effects and interactions) design into a power analysis (effect size $f = .05$, power = .80, groups = 18, number of predictors = 7 [3 main effects, 3 two-way interactions, 1 three-way interaction], response variables = 1). In order to achieve this result, an initial target sample of 622 was collected due to concerns regarding attrition and for data quality (Zhou and Fishback, 2016). Of these participants, study attrition from Time one to Time two resulted in the loss of 96 people (15.4% of the sample), and further data cleaning resulted in the removal of participants who failed attention checks (14 people), those who failed to adequately comply with training at Time 1 (48 people; more detail below), time on page (less than two minutes as in Study one; 43 people), and CAPTCHA verification (3 responses) led to the total removal of 204 cases, resulting in 418 final cases used in final analyses, which surpassed the power analysis goal of 295.

Procedure

In Study two, a second Mechanical Turk study was built on the design of Study one to evaluate the effectiveness of two diversity trainings in reducing discrimination in interviews. Participants in a control group received internet safety training. The randomized experimental design for Study two was a 2 x 3 x 3, with two levels of interviewee race (White, Arab), three levels of interviewee performance (low, moderate, high), and three levels of participants’
participation in diversity training (control/no diversity training, goal-setting diversity training, perspective-taking diversity training). This study included two time points. At Time one, participants were randomly assigned to one of the three training conditions (control, perspective taking, goal setting) and completed the relevant training course (see details below). Each training course was approximately 45 minutes. All participants were invited back online for Time two at the completion of two weeks, and were then exposed to one condition from the 2 x 3 design utilized in Study one that crossed interviewee race with competence level. Participants provided three ratings, one rating per interview question, using the same materials as Study one. Finally, additional measures such as manipulation checks and demographic variables were collected. This time lag was necessary to allow time for participants to practice the skills they learned during training and to work towards the goals they set prior to the experimental manipulation of the interviews in order to evaluate the effectiveness of the training protocols on the suppression and/or justification of discrimination in interview ratings (Bezrukova, 2016).

Training materials for the two diversity training conditions were modeled on past research efforts with perspective taking protocols (Galinsky and Ku, 2004; Galinsky and Moskowitz, 2000; Macrae et al., 1994) and goal-setting protocols (Lindsey, King, and Hebl, 2015; Madera, King, and Hebl, 2013), with more explicit detail for training content provided in the Materials section below. In order to provide a valid comparison group for the two diversity training conditions we included a no-treatment control condition that completed an irrelevant training on internet safety. This course was a shortened version of the free, online internet safety course developed by Goodwill Community Foundation, Inc., a developer of a large selection of online learning programs (http://www.gcflearnfree.org/internetsafety). There were knowledge check questions throughout with a take-home activity of completing three recommended actions.
to make their internet presence safer. See Appendix B for more information. This procedure allowed for similar resource demands between control and experimental groups over the two-week time lag.

Materials

*Online Diversity Training Content.* In order to evaluate the effects of diversity training on the justification and suppression of race-based discrimination effect in employment interviews, training content relating to (a) individuals taking the perspectives of others and (b) setting goals around being more inclusive with diverse others were utilized. Participants in both training settings watched a set of four short videos that discuss tragedies that are borne out of prejudice, positive experiences from diverse interactions, and community leaders reflecting on their diverse experiences (one 8-minute video, one 7-minute video, and two 3-minute videos; total run time = 21 minutes). Time one of Study two had a mean completion time of 44 minutes for the internet safety condition, 50 minutes for the perspective taking condition, and 43 minutes for the goal setting condition.

The perspective taking condition had participants answer short open text response items prompting reflection and taking the perspectives of targets in each video segment (e.g. *Imagine you are one of the high school students who immigrated to the U.S. – in 1-2 paragraphs describe what your experience might be like if you moved to a new country as a High School student.*). This procedure is an online adaptation to Galinsky and Moskowitz’s (2000) intervention where individuals are presented with scenarios and are asked to write about the experience of the target individual in the scenario (which was originally sourced from Macrae et al., 1994).

Participants in the goal setting condition filled out two goal setting worksheets related to diversity and inclusion related goals that they set for themselves (see Appendix B for materials).
This approach is an online adaptation from Madera, King, and Hebl (2013) in which students on a college campus shared personal stories in which they faced some challenge related to their identity while a student at the university. Students were subsequently provided with goal setting materials and walked through aspects of successful goals (e.g. S.M.A.R.T) (e.g. *Does this goal answer the question of much/ how many/ or how often.* for “Measurable”). In order to account for this additional instruction my goal setting materials have helpful tips to good goal setting on the same form they completed for the study. At the completion of the training, participants in both groups were encouraged to take advantage of the two weeks between experiment sessions to implement what they had learned or what goals they had set.

Participants in the internet safety condition viewed videos, read text, and participated in knowledge check questions around creating safe passwords, shopping safely online, and avoiding scams. At the end of the training, participants were asked to pick three action items from a list to complete over a two-week period (e.g. make my online passwords more secure).

**Measures**

*Diversity Training Compliance.* In order to measure compliance with the training manipulations, when participants returned for Time two, I asked for some verification of their participation depending on what training they saw (Perspective taking condition: asked them to re-iterate their self-reflections; Goal setting condition: asked to upload the two completed goal setting forms; Control condition: asked which three internet safety activities they accomplished). Forty-eight participants who failed this compliance check or indicated that they did not do the activities were excluded from analyses.

*Utility Reactions.* To examine the success of the training, items were adapted from Morgan and Casper’s (2000) published scale of training reactions, that have been subsequently
used in diversity training evaluations (Holladay and Quinones, 2005; Velada and Caetano, 2000). Specifically, utility reactions have been shown to predict whether a participant will engage and learn from the content (Aguinis and Kraiger, 2009; Bezrukova et al., 2016). An example item is, “How important is the content covered in this course?” (1 = “not at all important”; 5 = “extremely important”). Score reliabilities were good (α = .92).

**Random Responding.** In order to guard against random responding and poor data quality, a random responding question was asked after each diversity training video was played. An example of such a question is, “Select “6” from the dropdown”. Participants that failed two or more of these four attention checks were excluded from analyses.

**Demographic Questions.** Additional information was gathered from participants to ensure the sample was representative of the intended population. Demographic questions included variables such as gender, race, and ethnicity, current country of residence (Sample restricted within MTurk to U.S only), and if they were a native English speaker (97% Native English speaker; 3% English as a second language). Additionally, questions relevant to their current work situation were included such as the number of hours worked per week, the number of interviews performed as a rater (M = 7.2, SD = 5.0), the number of interviews performed as an applicant, and industry of work.

**Analysis Strategy**

The analysis strategy used for Study two was similar to that of Study one with the addition of a third independent variable for a 2 x 3 x 3 fixed effects ANOVA for race (Arab, White) x interview competence level (low, moderate, high) x diversity training condition (control, perspective taking, goal setting) with interview scores as the dependent variable. However, Hypotheses 2a-3b required partitioning of conditions through a split file command, so
each required a distinct model for testing. Hypothesis 1 was re-evaluated on the omnibus 2x3x3 model to test for a two-way interaction of race with interview competence level. Hypothesis 2 evaluated the addition of diversity training condition to compare the race difference (Arab versus White) on interview scores overall, with Hypothesis 2a comparing both diversity training conditions to the control group, and Hypothesis 2b comparing the two diversity training conditions. Hypotheses 3a and 3b focused on race differences between conditions as well, but only at low levels of interview competence where justification factors are most present. This was evaluated through both an ANOVA model and subsequently comparing the effect differences at low interview competence between White and Arab conditions for the various training groups (i.e., 3a: Training/ No Training, 3b: Perspective Taking/Goal Setting). ANOVA normality assumptions were tested through the Shapiro-Wilk test and the homogeneity of variance assumptions were tested through Levene’s test; no violations were found.

Study two Results

Descriptive statistics and intercorrelations of all study variables are presented in Table 2. Interview scores as well as warmth and competence ratings were all significantly correlated, suggesting that interviewees in the higher competence conditions were seen as both more competent and warm. Additionally, the counter-intuitive correlation from Study one between conservative political orientation with higher competence ratings and vice versa was observed again in Study two, though not for warmth ratings. Interview scores were also positively correlated with the interview condition manipulation, as would be expected, but also with utility reactions from the trainings, suggesting that people who saw their Time one training as more useful or important also tended to rate applicants at Time two higher than those who got less out of their Time one training. Utility reactions were also negatively correlated with training
condition suggesting that participants preferred the internet safety control condition over the diversity trainings.

Table 2

*Correlations Among and Descriptive Statistics for Key Study 2 Variables*

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interview Scores</td>
<td>3.18</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Warmth</td>
<td>3.62</td>
<td>1.00</td>
<td>.26**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence</td>
<td>3.72</td>
<td>0.91</td>
<td>.35**</td>
<td>.64**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Political Orientation</td>
<td>3.97</td>
<td>1.83</td>
<td>.05</td>
<td>-.08</td>
<td>-.11*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. White/Arab Condition</td>
<td>1.50</td>
<td>0.50</td>
<td>-.04</td>
<td>-.02</td>
<td>.08</td>
<td>-.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Training Condition</td>
<td>2.08</td>
<td>0.79</td>
<td>.02</td>
<td>-.09</td>
<td>-.01</td>
<td>.09</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Utility Reactions</td>
<td>3.57</td>
<td>1.14</td>
<td>.15**</td>
<td>.05</td>
<td>.09</td>
<td>.07</td>
<td>.11*</td>
<td>-.13**</td>
<td></td>
</tr>
</tbody>
</table>

Notes. N = 319. Political Orientation coding: 1= Far Right Conservative, 7= Far Left Liberal. For White/Arab condition: 1 = White condition, 2 = Arab Condition – correlations with this variable are point-biserial. Training Condition: 1 = Control, 2 = Goal Setting, 3 = Perspective taking.
* p < .05, ** p < .01.

Before partitioning the data to test the specific hypotheses, an omnibus ANOVA 2x3x3 model was calculated to examine relationships among key variables in Study two. The main effect of interview performance condition was replicated in Study two, $F(2, 400) = 7.11, p < .01, \eta^2_p = .03$. The main effects of both ethnicity condition, $F(1, 400) = 1.32, p = .25, \eta^2_p < .01$, and training condition, $F(2, 412) = 1.77, p = .17, \eta^2_p = .01$, were non-significant, suggesting that participants rated the interview responses for the Arab and White conditions similarly as well as those who participated in different training conditions. Additionally, in a direct replication of Hypothesis 1, I found no evidence of a significant two-way interaction effect of ethnicity and interview competence level, $F(2, 400) = 1.70, p = .18, \eta^2_p = .018$, suggesting that the similarity
in interview ratings between the Arab and White interviewee condition was consistent across the three levels of interview competence. The interaction between interview performance condition and training condition was also non-significant, \( F(4, 400) = 1.02, p = .40, \eta^2_p = .01 \), suggesting that diversity training did not differentially impact ratings of low, average, or high performing interviewers. Additionally, the interaction between ethnicity condition and training condition was non-significant, \( F(2, 400) = 1.43, p = .24, \eta^2_p = .01 \), suggesting that diversity training condition did not impact overall ratings for Arabs versus Whites. Finally, the three-way interaction between interview competence condition, ethnicity condition, and training condition was also non-significant, \( F(4, 400) = 1.96, p = .10, \eta^2_p = .02 \), suggesting that diversity training condition did not impact the difference between White and Arab interviewees in different ways across the three levels of interview competence. These ideas are further tested in more specific comparisons that align with my hypotheses below.

First, Hypothesis 2a was tested through a 2 (race) x 3 (interview competence) x 2 (diversity training versus control) ANOVA model. The main effect of the interview condition, \( F(2, 412) = 5.26, p < .01, \eta^2_p = .03 \) was significant, suggesting participants were differentiating between interview conditions. The main effect of ethnicity was not significant, \( F(1, 412) = 1.92, p = .17, \eta^2_p = .01 \), suggesting the resume manipulations did not significantly impact interview scores. The main effect of diversity training (versus control) condition was also not significant, \( F(1, 412) = 0.24, p = .63, \eta^2_p < .01 \), suggesting that interview ratings were not significantly impacted by participating in any form of diversity training relative to the control group. The three-way interaction did not produce a significant result, \( F(2, 406) = 0.13, p = .88, \eta^2_p < .01 \) nor did any of the preceding two-way interactions (interview condition x ethnicity: \( F(2, 406) = 1.59, p = .20, \eta^2_p = .01 \); interview condition x diversity training condition: \( F(2, 406) = 0.41, p = .66, \eta^2_p < .01 \).
$\eta^2_p < .01$; ethnicity x diversity training: $F(2, 406) = 1.56, p = .21, \eta^2_p < .01$. These results suggest that differences in participants’ interview ratings between the Arab and White conditions (i.e., discrimination) did not significantly differ as a function of participation in a diversity training versus control condition, and these non-significant patterns were consistent across all three levels of interview competence. Further, the presence or absence of diversity training in a participant’s experience did not significantly interact with their viewing of an Arab versus White resume nor the competence in the video recordings presented to them following training. Hypothesis 2a was not supported.

To illustrate these findings, Figure 4 shows strong homogeneity within interview competence conditions, which graphically shows the lack of effect both between control versus diversity training conditions as well as the White and Arab resume manipulations. Whereas, the omnibus effects across the interview performance conditions were not supported, the mean differences between conditions at low interview performance (e.g. when justification factors for discrimination are present) suggest that there may be more subjective judgment for raters in that condition. This was explored further in Hypotheses 3a and 3b.
Hypothesis 2b, which split up the diversity training conditions to compare the differential impact of goal setting training versus perspective taking training on interview ratings while omitting the control group, was tested through a 2 (race) x 3 (interview competence) x 2 (goal setting versus perspective taking) ANOVA. The main effect of the interview condition $F(2, 292) = 5.58$, $p < .01$, $\eta^2_p = .04$ was again significant, suggesting participants were differentiating between interview conditions. The main effect of ethnicity was again not significant $F(1, 292) = 0.034$, $p = .85$, $\eta^2_p < .01$, suggesting the resume manipulations did not significantly impact interview scores. The main effect of goal setting versus perspective taking training was also not significant $F(1, 292) = 3.12$, $p = .08$, $\eta^2_p = .01$, suggesting that interview ratings were not significantly impacted by participating in the goal setting diversity training versus the perspective taking diversity training. Although there were no significant two-way interactions (interview competence condition by ethnicity condition, $F(2, 292) = 0.74$, $p = .48$, $\eta^2_p = .01$; interview competence condition by training condition, $F(2, 292) = 1.54$, $p = .28$, $\eta^2_p = .01$;
ethnicity condition by training condition, $F(2, 292) = 1.21, p = .27, \eta^2_p < .01$). There was a significant three-way interaction effect, $F(2, 292) = 3.53, p = .03, \eta^2_p = .001$. These results suggest that one of the two diversity trainings results in less bias between the White and Arab conditions, but in a different manner for at least one of the three levels of interview performance (Kirk, 1995).

Figure 5 helps to illustrate these differences. First, for the perspective taking condition, the Arab applicant is rated higher on average than the White applicant for both the high and medium (though more negligibly) interview performance conditions. However, this trend reverses in the low interview competence condition where the White applicant is rated higher than the Arab (consistent with a justification effect). The goal setting condition, by comparison, presents a different pattern. At medium and high levels of interview performance, goal setting and perspective taking produce inverse results, with goal setting training producing a favorable impact for the white ethnicity, while perspective taking benefits the Arab condition. At the low level of interview competence those in the goal setting condition demonstrate similar ratings for the White and Arab applicant, reducing the justification effect. As mentioned, Hypotheses 3a and 3b aimed to explore how the presence of justification factors (low interview performance) impacts the relationships of diversity training and the ethnicity of the interviewee, so these distinct patterns at low levels of competence will are more formally tested below. However, due to a lack of a two-way interaction between diversity training and ethnicity for overall interview ratings, Hypothesis 2b was not supported as neither goal setting nor perspective taking were more effective at reducing discrimination overall (across all levels of performance).
Hypotheses 3a/b focused more specifically on the effect of diversity training on reducing the potential justification of discrimination when it was most expected—at low levels of interview competence only, with H3a comparing any diversity training to the control and H3b comparing the goal setting and perspective taking diversity trainings. Hypothesis 3a was evaluated with a 2 (race) x 2 (diversity training versus control) ANOVA with interview scores at low levels of competence as the dependent variable. The main effect of training was not significant, $F(1, 131) = 0.95, p = .76, \eta^2_p = .001$, suggesting that participants were not rating applicants differently based on receiving diversity training or not. The main effect of ethnicity was also not significant $F(1, 131) = 3.77, p = .054, \eta^2_p = .03$, suggesting that the resume manipulations to present the applicant as Arab or White did not significantly impact interview scores when justification factors were present. Further, the two-way interaction of diversity training by ethnicity was not significant $F(1, 131) = 1.03, p = .31, \eta^2_p = .01$, suggesting the differences between White and Arab applicants, or the extent of possible discrimination, did not
differ between the combined diversity training groups and the control condition. This lack of an interaction failed to support Hypothesis 3a.

Hypothesis 3b re-ran a similar ANOVA on the low level of interview competence only, but instead compared the two types of diversity training more specifically, testing main effects and the interaction of race (Arab, White) and diversity training type (goal setting, perspective taking). The main effect of diversity training type was not significant, $F(1, 94) = 0.03, p = .87, \eta_p^2 < .01$, suggesting that interview ratings at low levels of competence were not significantly impacted by participating in a goal setting training versus perspective taking diversity training. Further, the main effect of ethnicity was also not significant, $F(1, 94) = 0.80, p = .38, \eta_p^2 = .01$, conveying that the resume manipulations were not significantly impacting the observed interview ratings (i.e., a lack of discrimination in interview ratings between groups). Despite the three-way interaction reported above when all levels of interview competence were considered, the two-way interaction effect between ethnicity and diversity training at low levels of interview competence only also failed to produce a significant result, $F(1, 94) = 2.38, p = .13, \eta_p^2 = .03$. This suggests that the lack of ethnicity-based differences in interview ratings at low levels of competence did not differ between the goal setting and perspective taking diversity training conditions. In sum, Hypothesis 3b was not supported.

Due to the descriptive trends observed in Figures 3 and 4 as well as some of the trending statistical results in Study two, one might ask the question of the practical significance of these findings. Put another way, practitioners still need to make organizational decisions about diversity training and since this study’s aim was, in part, to help make those decisions, in Table 2 I present condition and effect size differences on interview scores between ethnicity and training conditions when justification factors are present. A Cohen’s $d$ effect size directly measures the
strength of a relationship between two variables in a sample population (Diener, 2010). In this case, a strong effect size difference between Arab and White conditions means that the effect of the resume manipulation was stronger for one group versus another (e.g. training vs. no training). Smaller effect sizes between Arab and White conditions for a given condition suggest that raters have suppressed their bias more. The results presented in Table 2 suggest that training may be an effective way to reduce rater bias when justification factors are present since the difference between White and Arab conditions was smaller for those who participated in any form of diversity training ($d = 0.16$) than for those who participated in no form of diversity training and showed larger preferences for White versus Arab applicants ($d = 0.65$), although as pointed out above for the test of H3a, these mean differences were not statistically different. Similarly, the goal setting training condition produced smaller effect size differences between the White and Arab low interview competence conditions that in fact slightly favored Arabs over Whites ($d = -0.13$), compared to the Perspective taking conditions ($d = 0.49$). However, as displayed in our H3b results above, these differences were not statistically significant. Nonetheless, for a practitioner, these directional results suggest that instituting a goal setting form of diversity training for those who conduct employment interviews may help to reduce rater bias.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>Perspective Taking</th>
<th>Goal Setting</th>
<th>Training</th>
<th>No Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>White interview mean</td>
<td>3.16</td>
<td>2.91</td>
<td>3.03</td>
<td>3.26</td>
</tr>
<tr>
<td>Arab interview mean</td>
<td>2.71</td>
<td>3.03</td>
<td>2.88</td>
<td>2.75</td>
</tr>
<tr>
<td>SD White</td>
<td>1.01</td>
<td>0.91</td>
<td>0.96</td>
<td>0.88</td>
</tr>
<tr>
<td>SD Arab</td>
<td>0.83</td>
<td>0.90</td>
<td>0.87</td>
<td>0.70</td>
</tr>
<tr>
<td>N White</td>
<td>25</td>
<td>25</td>
<td>50</td>
<td>18</td>
</tr>
<tr>
<td>N Arab</td>
<td>23</td>
<td>25</td>
<td>48</td>
<td>19</td>
</tr>
</tbody>
</table>
In line with Study one, warmth and competence data were also collected in Study two to examine not only outcome oriented results (interview ratings), but also to capture differences in how participants viewed Arab and White applicants. When warmth scores were entered in place of interview scores as a dependent variable in the omnibus 2x3x3 ANOVA model, the main effects of interview condition $F(2, 400) = 1.47, p = .23, \eta^2_p = .01$ and ethnicity condition $F(1, 400) = .08, p = .385, \eta^2_p < .01$ were non-significant. However, the main effect of training condition was significant $F(2, 400) = 3.46, p = .03, \eta^2_p = .02$, suggesting that warmth scores may be significantly differentiated by the training participants completed at Time one. Three Post-hoc t-tests for pairwise comparisons between training groups with a Bonferroni correction (significant $p$-value set at .017) suggested significant differences were only found between the goal setting ($M = 3.71, SD = 0.73$) and perspective taking condition ($M = 3.50, SD = 0.69$) $t(302) = 2.6, p = .011$. The control condition ($M = 3.65, SD = 0.59$) vs. goal setting condition produced a non-significant result of $t(270) = -0.73, p = .47$ as well as the control condition vs. the perspective taking condition with a result of $t(258) = 1.83, p = .07$. These results suggest that warmth scores in the two training conditions did not significantly deviate from those who saw the control training condition.

Continuing with the rest of the 2x3x3 ANOVA model for warmth scores, all two way interactions were non-significant, including interview performance condition by ethnicity condition, $F(2, 400) = 1.73, p = .18, \eta^2_p = .01$, interview performance condition by training condition, $F(4, 400) = 1.37, p = .24, \eta^2_p = .01$, and ethnicity condition by training condition,
$F(2, 400) = 0.76, p = .47, \eta_p^2 < .01$. The three-way interaction was also not significant, $F(4, 400) = 1.62, p = .17, \eta_p^2 = .02$. In sum, these results are consistent with the omnibus findings when interview scores were the dependent variable.

When competence scores were entered in place of interview scores as a dependent variable in the omnibus 2x3x3 ANOVA model, the main effects of interview condition $F(2, 400) = 0.69, p = .50, \eta_p^2 < .01$, ethnicity condition $F(1, 400) = 0.84, p = .36, \eta_p^2 < .01$, and training condition $F(2, 400) = 1.26, p = .29, \eta_p^2 = .01$, were all non-significant, suggesting that none of the manipulations influenced participants’ competence views of the applicants. Two of the two-way interactions, interview performance condition by ethnicity condition, $F(2, 400) = 0.71, p = .49, \eta_p^2 < .01$, as well as the interview performance condition by training condition, $F(4, 400) = 1.03, p = .39, \eta_p^2 = .01$, were non-significant. However, the ethnicity condition by training condition interaction was significant $F(2, 400) = 6.60, p < .01, \eta_p^2 = .03$, suggesting that competence judgments may significantly differ when comparing the Arab applicant and white applicant for at least one of the training conditions. However, three post-hoc one-way ANOVAs, Bonferonni corrected (significant $p$-value set to .017) comparing ethnicity condition within each level of training condition found no statistically significant differences between ethnicity conditions. The control training condition produced a one-way effect between ethnicity conditions on competence scores of $F(1, 112) = 4.37, p = 0.039$. The goal-setting training condition produced a one-way effect $F(1, 156) = 4.79, p = 0.030$. The perspective taking training condition produced a one-way effect of $F(1, 144) = 3.98, p = 0.048$.

Continuing in the 2x3x3 model for competence scores, the three-way interaction was also significant $F(4, 400) = 2.53, p = .04, \eta_p^2 = .03$, suggesting that the difference between Arab and White competence ratings differed between training groups to different degrees for at least one
of the three interview performance levels. However, as shown in figure 6, competence ratings do not always share the anticipated positive linear relationship with interview performance. The control condition for both ethnicities had a negative linear relationship with interview performance. Additionally, within the Arab ethnic condition, both perspective taking and goal setting had the highest competence scores at average levels of interview performance. The goal setting condition where participants were also exposed to the white ethnic manipulation was the only condition that shared a positive linear relationship between competence ratings and interview performance.

![Figure 6: Results for Competence scores across all conditions. Error bars represent standard error of the mean.]

**Study two Discussion**

Although Study one did not detect the hypothesized relationships between interview competence and ethnicity of the applicant, Study two still provides important findings in regards to how and which kind of diversity training may impact the justification of discrimination against Arab applicants who are less competent. Hypothesis 2a, that diversity training would have a
positive impact on interview scores in regards to rating distortions for Arab applicants was not 
found in Study two. Given that in Study two the ethnicity main effect was also non-significant, 
this is not a surprising finding. If the bias that the trainings are trying to reduce is statistically 
negligible, then the availability for impact of the trainings is also reduced. Similarly, for 
Hypothesis 2b that the differential impact of a goal setting diversity training versus a perspective 
taking training on the interview performance by ethnicity relationship were non-meaningful 
likely for this same reason, that I failed to find strong statistical bias that needed eliminating. The 
general consistency that warmth and competence ratings shared with the hypothesis tests when 
evaluated as dependent variables strengthens the lack of statistical bias in the study in that the 
evaluative pathways did not significantly deviate from the observed final results (interview 
scores). This, of course, can be construed as a positive finding since Arabs and Whites may not 
be as susceptible to differential interview ratings, even when the justification for discrimination 
may seem more warranted because interviewers have another reason to explain the potential for 
harsher evaluations of Arabs who are viewed as less competent and warm (Neumark, 2018). 
Indeed, the significant 3-way interaction between ethnicity condition, interview performance, 
and training condition observed while testing hypothesis H2b is suggestive directionally that 
there are different patterns of bias that vary as a function of interview performance levels, with 
the goal setting intervention more effectively reducing such bias than the perspective taking 
condition.

When it comes to examining how diversity training impacts people suppressing their rater 
bias when both Arab and White applicants are performing poorly on the interview, Study two 
failed to find statistically significant effects for training reducing bias. However, effect sizes 
comparing mean group differences highlight some positive trends that researchers should
examine further. At low levels of interview competence, effect size differences between Arab and White applicants were smaller for those that received training ($d = 0.16$) compared to people who did not receive any form of diversity training ($d = 0.65$) suggesting that diversity training may yet have efficacy in reducing bias in interview ratings when justification factors are present. Further, when it comes to comparing the relative efficacy of perspective taking and goal setting diversity trainings, the provisional evidence of effect size trends in Study two point to goal setting as the potentially more efficacious avenue in order to reduce the impact of bias even when justification factors are present ($d = -0.13$ for goal setting versus $d = 0.49$ for perspective taking). These conclusions are, of course, tempered by the fact that these mean group differences were not statistically significant—a fact that may be due to the relative weakness of the manipulation employed (resume with name and background information altered) as opposed to the possibility of stronger manipulations to highlight ethnicity differences between the White and Arab applicant conditions. Future studies should look to test these directional differences further with different ethnic manipulations.
Chapter 4

Discussion and Conclusion

General Discussion

This investigation first, examined the expression of modern bias through the lens of the justification-suppression model (Crandall and Eshleman, 2003) in order to detect and explain possible discrimination Arabs may experience during employment interviews. Subsequently, this investigation also looked to examine the impact diversity trainings would have in amending possible discrimination against Arabs. Study one did not find support for the proposed justification of discrimination against the Arab, as opposed to White, interviewee as a function of how well the candidate performed. Instead, suppression of discrimination or, perhaps a lack of discrimination, was evident at all levels of interview performance when comparing the Arab and White conditions. Further, neither ethnicity nor interview performance level explained a significant amount of variance in warmth or competence scores. Study two replicates and extends these findings by providing insights regarding how and which kind of diversity training may better reduce the justification of discrimination against Arab applicants who are less competent. Overall, however, participation in diversity training did not significantly impact observed interview scores. Given that differences between White and Arab scores were non-significant in both studies, these null results are not surprising since the opportunity to reduce biased ratings did not present itself in either of these studies. In other words, it is difficult to test ideas to eliminate bias when you cannot detect bias in the first place. When examining the differential impact of a goal setting diversity training versus a perspective taking training there was some evidence that the differences between White and Arab ratings depended on the level of interview performance. However, when examining how training influence ratings for poor
performing candidates, more specifically, where the conditions to justify discrimination against stigmatized groups such as Arabs are the strongest (Crandall and Eshleman, 2003). Study two failed to find statistically significant evidence of diversity training reducing bias. However, effect sizes comparing mean group differences and a significant three-way interaction highlight some positive indications that researchers should examine further that are in line with my hypotheses. In particular, these trends suggest that at low levels of interview competence, diversity training may have efficacy in reducing bias in interview ratings, arguably because this is when justification factors are present and interviewers may benefit the most from intervention to reduce harsher punishment of an Arab poor performer relative to a White poor performer.

Further, when it comes to comparing the relative efficacy of perspective taking and goal setting diversity trainings, these same findings point to goal setting as the potentially more efficacious avenue in order to reduce the impact of bias even when justification factors are present. However, due to the relative weakness of the manipulation employed, further research is needed to confirm these trends.

Study Limitations

As mentioned previously there are several limitations to the investigation. First, as was discussed extensively in the Study one discussion, the ethnic manipulation was relatively weak through resume manipulation, so future studies should look for stronger indicators of Arab status. It is possible that a stronger manipulation of interviewee ethnicity (e.g., in dress and grooming, accent, etc.) may have led to more consistent levels of discrimination as seen in past research (King and Ahmad, 2010). Additionally, the participants in these studies were providing low-stakes interview scores, and as previous research has shown, high-stakes employment settings versus low-stakes lab studies has the potential to produce significantly different results (Derous,
Ryan, and Nguyen, 2012). Future studies may find value in utilizing archival data from high-stakes settings, where justification-suppression type processing may be more prominent. Additionally, participants were recruited through MTurk and whereas this sample is more diverse and representative of a general population compared to a university sample, choosing to go through Mturk introduces selectivity bias in who could participate in the study in that those that participated are more likely to be younger, less religious, and more educated than the general population (Paolacci, and Chandler, 2014). Additionally, the diversity trainings presented in Study two may have not have been effective enough to produce the desired impact. To expand, some diversity training videos may have been too short to engage participants thoroughly and the activities for the respective goal setting and perspective taking conditions may not have been engaging enough especially in an online environment for learning or transfer to occur (Carter, Onyeador, and Lewis Jr, 2020). Future studies should look to deliver diversity trainings that allow for aspects error management, practice, asking questions, as well as exploring an array of modalities to support learning transfer of the content. It may have also been the case that participants were clued into the research aims of the two studies as evidenced by negative relationship between political orientation and warmth and competence scores. Participants may have been clued in and choose to answer in socially desirable ways rather than in a genuine fashion. Future studies should look to investigate paradigms with less demand characteristics as well as include direct measures of social desirability. Another limitation in utilizing the resume manipulations is that by changing the locations of the applicant’s education and work history, not only is Arab status manipulated but also immigrant status, which is a conflation in the manipulation. Participants may have reacted to the fact that either the applicant was an immigrant or that they presented as an Arab. Also, participants may have struggled to interpret
the signaling of education and organizational institutions from another country in which they have no frame of reference. Future studies should look to tease apart the effect of Arab status and immigrant status through both resume manipulations as well as other stimuli that can control for the effect of one while allowing for the other to vary. Finally, the interview competence manipulation, while statistically significant in hypothesis testing, did not produce very strong differentiation between performance conditions. It may have been the case that the video scripts of the applicant did not produce enough signals for applicants to use the BARs anchors appropriately. Future studies should look to either gather high-stakes organizational data or create vignettes with stronger differentiation between performance conditions. Despite these limitations there are still important implications and contributions to be gleaned from this investigation, as well as avenues for future research.

Implications

There are several key contributions this investigation makes to the Organizational Diversity and Personnel Selection literature bases. First, this investigation answers calls for research from several studies. Levashina et al. (2014) and Derous, Buijsrogge, and Duyck (2016) asked for more consistency and structure in experiments to allow for better minority vs. majority comparisons, thus reducing the number of possible explanations for observed differences. Additionally, Derous, Buijsrogge, and Duyck (2016) called for research to examine interviewer experience and the magnitude of their biases, which this study can examine. Ruggs et al. (2013) chastised the I/O community for largely ignoring the present day issues with diversity in the workplace and more specifically extended a call for research to investigate the differential rates of mistreatment for minorities based on important moderating variables. Therefore, this investigation answers these calls by investigating applicant competence as an explanatory
variable for subgroup differences, evaluate the magnitude of interviewer bias through a new lens, and examine differential treatment to minorities vs. majority applicants within the same levels of competence. Moreover, there are many modern -isms (ex. sexism, heterosexism, ageism) that impact organizational effectiveness (Jones et al., 2017). This investigation contributes to understandings in the modern expression and impact of implicit biases in the interview setting at large.

An additional contribution of the current study is that although previous research using the justification-suppression framework has supplied strategies for stigmatized victims to employ to reduce adverse treatment, few studies have focused on what the transgressor can do to reduce their disparaging behaviors (Brelan, Seitz, Treadway, Lovelace, and Gazdag, 2017; Hebl et al., 2007; Randall et al., 2017). This investigation placed the burden of intervention on the organization to focus on how organizations can train the right behaviors and attitudes for their interviewers. EEOC reports (see Figure 1) demonstrate that organizations have been unable to move the needle on disparate treatment. Through exploring the efficacy of diversity trainings’ ability to impact bias, practitioners have some directional evidence that diversity trainings backed by a goal setting methodology may be more efficacious than those backed by a perspective taking methodology, although more work is needed to confirm this directionality.

Diversity training theory also benefits from this investigation by pitting goal-setting and perspective-taking against one another as rival approaches to reducing racial/ethnic discrimination in the interview context. The literature thus far has suggested that stereotype discrediting may not be as viable an approach to improving diversity related attitudes and behaviors, but there is little evidence to suggest which approach to diversity training is better (Hebl and King, 2013). Further theory testing is still necessary, but the directional evidence from
this investigation points to a goal-setting theory framework over a perspective taking framework lending itself to more learning transfer when trying to reduce bias between Arab and white applicant interview scores.

Finally, this investigation approaches these research questions from the perspective of trying to bridge the gap between science and practice. The EEOC has seen an increase in the number of their race related complaints, however the literature (Huffcutt and Roth, 1998; Levashina et al., 2014; Posthuma, Morgeson, and Campion, 2002) is detecting smaller race differences on mean interview scores. Those two ideas become harder for one person to hold unless they have a reason why. This investigation provides initial evidence for an explanation of those seemingly disparate findings (i.e., considering both suppression and justification factors) by examining a range of interview performance: from poor to excellent.

Theoretical Contributions and Future Directions

The experimental design choices for these studies offer theoretical contributions to the workplace diversity literature. By experimentally controlling for other proximal characteristics except for the resume manipulations, these studies remove explanations due to gender, religion, difficulties understanding accents, etc. that are often confounded in diversity research. However, as addressed above, these gains in internal validity came at the cost of construct validity strength as Arabs rarely present with only one or even few stigmatizing features. Further research and theory should consider how to capture and partial the often multi-faceted nature of applicant’s identities and features and how they come together to interact in employment settings. Additionally, Derous, Buijsrogge, and Duyck (2016) called for research to examine interviewer experience and the magnitude of their biases, which was examined. Overall, the effect of ethnic bias was small and non-meaningful at times. This small effect is consistent with the majority of
research on subtle manifestations of prejudice (e.g., Hebl et al., 2002; King et al., 2006; Singletary and Hebl, 2009) and suggests that social norms and legislation are restricting or suppressing overt forms of discrimination. It is important to note, however, that even small amounts of bias in employment decisions can accumulate over time and create substantial differences in the careers of advantaged and disadvantaged group members (Martell, Lane, and Emrich, 1996).

One of the key gaps in the diversity training literature was tying training to the subsequent outcomes beyond attitudes and behavior. Applying this focus addresses a key limitation in the diversity training space (Bezrukova, Jehn, and Spell, 2012; Kraiger, Ford, and Salas, 1993; Kulik and Roberson, 2008; Nittrouer, Hebl, and Oswald, 2016). This investigation adds to the literature through showing that when justification factors are present, that individuals that experience a diversity training may suppress their bias more than individuals who do not experience training, but this should be explored in further in future studies. Further, exploration in Study two nods to the findings of Bezrukova et al.’s 2016 meta-analysis in that they found goal setting diversity training is more efficacious than perspective taking diversity trainings. It may be the case that targeting behavior change more directly through setting goals rather than through an attitudinal pathway creates a stronger impact on outcomes.

One potential future direction for this research is to test potential interactions between the resume characteristics and differing stigmatizing features of the applicants. For example, we might expect that a visibly Muslim Arab that studied in the United States might face differing prejudice than an Arab who does not present as Muslim and studied in an Arab country. Another direction for future research is to examine the impact of racial/ethnic bias within different steps of the recruitment process. Applicants and organizations are making many choices in a
recruitment and selection process that have a possibility for bias from an organization deciding on the minimum qualifications for a role, to an applicant deciding and accepting an offer. The impact of modern racial/ethnic bias on advertising open roles, demographics of interviewers, stereotype congruity between the applicant and role are all possible future directions that help researchers to understand where implicit and covert forms of bias exist in organizational systems.

Practical Contributions and Future Directions

This study serves the practitioner audience by supplying evidence based avenues to target biased ratings. One potential direction for researchers involves the complexities and intersectionalities around presenting with an Arab status. This investigation manipulated Arab status through only a resume manipulation, which is but one way individuals signal their Arab status. However, many Arabs present with intersectionalities such as religious and gendered stigmas. In a field study with experienced recruiters, Derous, Ryan, and Serlie (2015) found that Arab woman were rated more favorably than Arab men, especially for roles that require high levels of customer contact. This points to the complex and nuanced nature of the Arab stigmatizing experience and future studies should look to investigate and remediate these intersectional biases. This investigation also adds practical contributions to the use of the JSM. Previous research using the justification-suppression framework has supplied strategies for stigmatized victims to employ to reduce adverse treatment, yet few studies have focused on what organizations can legally or ethically do to address such issues (Breland, Seitz, Treadway, Lovelace, and Gazdag, 2017; Hebl et al., 2007; Randall et al., 2017). First, as demonstrated by the indicators of practical, though not statistical significance, from Study two, evidence-based diversity training, whether in the form of perspective taking or goal setting, may be better than no training at all. I say this cautiously since direct comparisons between the control and diversity
training groups in this study were not significantly different; however, with the majority of the diversity training literature pointing to beneficial outcomes that should result in reduced discrimination (Bezrukova et al., 2016), it is clear that efforts to encourage diversity and inclusion, while simultaneously eliminating illegal discrimination based on applicants’ protected class status, should lead to improved organizational outcomes. Second, if practitioners can only choose one, then a goal setting approach may likely lead to superior outcomes compared to perspective taking—a view that is consistent with past research (Bezrukova et al., 2016).

Conclusion

Stigmatized applicants, such as Arabs applying for jobs in America, may experience unfair treatment and evaluation in interview settings, putting them at a disproportionate disadvantage to achieve gainful employment. Although the results presented here did not find strong evidence of bias favoring White over Arab interviewees, theoretical arguments and mixed empirical evidence identifies those stigmatized individuals who perform more poorly as most at risk to being the targets of race-based discrimination compared to those who perform better. However, some limited evidence points to diversity training, particularly goal-setting diversity training as a potential avenue to limit the amount of discrimination poor-performing Arabs may receive relative to their poor-performing White counterparts. Given the fraught nature of global intergroup relations and the increasing diversity of the American workforce, a more nuanced view that overlays stigmatized identities with factors like competence that may justify engagement in discriminatory behavior, provides important insights into how and when to expect discrimination to occur. Although not always easily addressed, I/O Psychologists must concern themselves with solving these challenges or be faced with questions like “who does our work serve?”.
References


Bushman, B. J., and 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.


Dipboye, R. L. (1997). Organizational barriers to implementing a rational model of training. In M. A. Quiñones, and A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research; training for a rapidly changing workplace: Applications of psychological research* (pp. 31-60, Chapter xii, 345 Pages) American Psychological Association, Washington, DC. doi: [http://dx.doi.org/10.1037/10260-002](http://dx.doi.org/10.1037/10260-002)


EEOC. 2016a. *EEOC Releases Fiscal Year 2016 Enforcement and Litigation Data*

EEOC. 2017. *Charge Statistics (Charges filed with EEOC) FY 1997 Through FY 2017*


Huffcutt, A. I., and Woehr, D. J. (1999). Further analysis of employment interview
http://dx.doi.org/10.1002/(SICI)1099-1379(199907)20:4<549::AID-JOB921>3.0.CO;2-Q


Corrections for range restriction in structured interview ethnic group differences: The values may be larger than researchers thought. *Journal of Applied Psychology*, 87(2), 369.


Taggar, S., and Kuron, L. K. (2016). The toll of perceived injustice on job search self-


Appendix A: Interview Questionnaire

All materials can be found at https://osf.io/g6ku4/files/

(Adapted from Source: State of Arizona)

Interview procedure

Thank you for taking part in this study. You are about to view two pre-recorded interviews of candidates interviewing for a sales role. You will be asked to rate each candidate’s response to each question (1 being the worst response and 5 being the best response). In order to assist you with this process, you will find examples of responses at certain numbers on the rating scale. These examples are called behavioral anchors and are meant to provide you with more information about the standard to compare candidates against. It is not likely that you will hear the candidates give the exact the same answer as the behavioral anchor provided, rather use the examples to find similarities with the responses when deciding what rating to give to a candidate’s response.

Additionally, below you will find a list of tasks that the candidate will be expected to perform if they were hired for the role. Please review these tasks to familiarize yourself with the role before moving on to the recorded interviews.

Lists of tasks salesperson will be expected to perform –

- Greet customers and ascertain what each customer wants or needs.
- Recommend, select, and help locate or obtain merchandise based on customer needs and desires.
• Compute sales prices, total purchases, and receive and process cash or credit payment.

• Prepare merchandise for purchase or rental.

• Answer questions regarding the store and its merchandise.

• Maintain knowledge of current sales and promotions, policies regarding payment and exchanges, and security practices.

• Demonstrate use or operation of merchandise.

• Describe merchandise and explain use, operation, and care of merchandise to customers.

• Ticket, arrange, and display merchandise to promote sales.

• Inventory stock and requisition new stock.

• Exchange merchandise for customers and accept returns.
Interview questions

**Question 1:** Describe a situation where you found yourself working on a project with a co-worker you have never met before. How would you make sure the project was done well?

Rating Scale:

(high answer) Meets with person 1-1, introduce each other to establish a working relationship and delegate tasks to one another. Set up frequent check-ins on progress.

(average answer) Sets up a single meeting, acts professionally but sets up a mechanism to maintain contact. May discuss delegating work between one another.

(low answer) Never meets with the person. Just tries to get their part of the project done. Displays non-professional behavior.

**Question 2:** Suppose you had many projects with rigid deadlines, but your manager kept requesting more paperwork, which you felt was unnecessary. This extra work was going to cause you to miss your deadlines. What would you do?

Rating Scale:

(high answer) Present the conflict to the manager. Suggest and discuss alternatives. Establish a mutually acceptable plan of action. Communicate frequently with the manager.

(average answer) Tell the manager about the problem. Request assistance. Request extension of deadline.
(low answer) Do the best I can. If I miss the deadline then it’s my manager’s fault.

**Question 3: Suppose you were going to run an important business meeting with your manager and leadership because they want to learn more about your recent major successes in your sales job. How would you prepare for the meeting?**

Rating Scale:

(high answer) I would catalog how I organize my work, look up my past sales numbers, my productivity metrics and customer feedback. Then, I would practice explaining how these data points exhibit the behaviors and attitudes I have engaged in to be successful.

(average answer) I would ask my manager what the meeting is about and take a few minutes to write down what I think lead to my recent successes.

(low answer) I would just show up to the meeting and answer my leadership’s questions.
Ethnicity (Manipulation check)

What ethnicity do you think the candidate most likely was?

- White.
- Hispanic or Latino.
- Black or African American.
- Native American or American Indian.
- Asian / Pacific Islander.
- Arab

Warmth and Competence Questions

Competence (1 not at all to 5 extremely)

As viewed by society, how competent are members of this ethnic group?

As viewed by society, how confident are members of this ethnic group?

As viewed by society, how independent are members of this ethnic group?

As viewed by society, how competitive are members of this ethnic group?

As viewed by society, how intelligent are members of this ethnic group?

Warmth (1 not at all to 5 extremely)

As viewed by society, how tolerant are members of this ethnic group?

As viewed by society, how warm are members of this ethnic group?
As viewed by society, how good natured are members of this ethnic group?

As viewed by society, how sincere are members of this ethnic group?
Appendix B: Diversity Training Materials (Study 2)

The participants who participate in study 2 went through a longitudinal design. In time 1, participants who are in the goal setting, as well as the perspective taking diversity training conditions received

1. High school students meeting each other and taking perspectives from each other:  
   https://www.teachertube.com/video/newcomers-254967

2. Patchogue, New York Mayor Paul Pontieri reflects on his family's history of immigration, his love of his hometown, and how his life has influenced his policy of inclusion for all Patchogue residents.  
   https://www.youtube.com/watch?v=jQqGPDYU5TAandfeature=youtu.be

3. As the Sikh community in Oak Creek, Wisconsin prepares for Sunday prayers, a deadly hate attack shatters their lives, but not their resilience. After six worshipers are killed by a white supremacist, the local community finds inspiration in the Sikh tradition of forgiveness and faith. Lieutenant Murphy, shot 15 times in the attack, joins the mayor and police chief as they forge new bonds with the Sikh community. Young temple members, still grieving, emerge as leaders in the quest to end the violence. In the year following the tragedy, thousands gather for vigils and community events to honor the victims and seek connection. Together, a community rocked by hate is awakened and transformed by the Sikh spirit of relentless optimism.  
   https://www.youtube.com/watch?v=PJbbw3b_ogsandfeature=youtu.be

4. Every January, Not In Our Town honors Dr. Martin Luther King’s legacy by sharing the real life stories of people who are applying Dr. King’s principles today. Though the political landscape has changed since the Civil Rights era, his dream that the United
States would fulfill its promise of equality has yet to become reality. But Dr. King’s work proves that change is indeed possible in this country. [https://youtu.be/SG7j1Ko7PPw](https://youtu.be/SG7j1Ko7PPw)
Perspective taking Materials

1. Video 1
   a. Imagine you are one of the high school students who immigrated to the U.S. – in 1-2 paragraphs describe what your experience might be like if you moved to a new country as a High School student.
   b. What new perspectives do you think the American students gained from interacting with their foreign counterparts?

2. Video 2
   a. How has Mayor Paul Pontieri’s experiences influenced his policies for Patchogue Residents?
   b. If you were the mayor of the current city you reside in, what polices would you implement in order to promote inclusion among your residents?

3. Video 3
   a. What kind of emotions do you think the family members of the six worshipers who were killed experienced and continue to experience?
   b. After viewing the community’s response to the hate attack, what do you think are the benefits of their response? What would you do differently if you were there?

4. Video 4
   a. What of Dr. King’s principles, could you apply today? Why?
   b. How do you think Dr. King would be trying to bring his dream to realization if he were alive today?

Goal Setting Materials

Participants were asked to fill out two forms, one for each goal.
Internet safety materials

(From: https://edu.gcfglobal.org/en/internetsafety/) All GCFGlobal.org® content is available for free at edu.gcfglobal.org.

Introduction
1. **Introduction to Internet Safety** - Get an introduction to Internet safety by understanding the risks and learning how to avoid them.

Staying Safe Online

2. **Creating Strong Passwords** - Use these tips to create a strong, safe password.

3. **Your Browser's Security Features print** - Better understand your browser's security features and how they work.

4. **Avoiding Spam and Phishing print** - Use these tips so you can avoid email-based spam and phishing scams.

5. **How to Avoid Malware print** - Learn how to avoid malware and remove it from your computer.

6. **Safe Online Shopping print** - Use these strategies for a safe online shopping experience.

From the list below – please pick 3 of the following activities to complete within the next two weeks – from ([https://securingtomorrow.mcafee.com/consumer/consumer-threat-notices/10-tips-stay-safe-online/](https://securingtomorrow.mcafee.com/consumer/consumer-threat-notices/10-tips-stay-safe-online/))

1. **Create Complex Passwords.** We know you’ve heard it before, but creating strong, unique passwords for all your critical accounts really is the best way to keep your personal and financial information safe. This is especially true in the era of widespread corporate hacks, where one database breach can reveal tens of thousands of user passwords. If you reuse your passwords, a hacker can take the leaked data from one attack and use it to login to your other accounts. Our best advice: use a password manager to help you store and create strong passwords for all of your accounts.
Then, check to see if your online accounts offer multi-factor authentication. This is when multiple pieces of information are required to verify your identity. So, to log into an account you may need to enter a code that is sent to your phone, as well as your password and passphrase.

2. **Boost Your Network Security.** Now that your logins are safer, make sure that your connections are secure. When at home or work, you probably use a password-protected router that encrypts your data. But, when you’re on the road, you might be tempted to use free, public Wi-Fi. The problem with public Wi-Fi is that it is often unsecured. This means it’s relatively easy for a hacker to access your device or information. That’s why you should consider investing in a Virtual Private Network (VPN). A VPN is a piece of software that creates a secure connection over the internet, so you can safely connect from anywhere.

6. **Protect Your Mobile Life.** Our mobile devices can be just as vulnerable to online threats as our laptops. In fact, mobile devices face new risks, such as risky apps and dangerous links sent by text message. Be careful where you click, don’t respond to messages from strangers, and only download apps from official app stores after reading other users’ reviews first. Make sure that your security software is enabled on your mobile, just like your computers and other devices.

7. **Practice Safe Surfing and Shopping.** When shopping online, or visiting websites for online banking or other sensitive transactions, always make sure that the site’s address starts with “https”, instead of just “http”, and has a padlock icon in the URL field. This indicates that the website is secure and uses encryption to scramble your data so it can’t be intercepted by others. Also, be on the lookout for websites that have misspellings or bad grammar in their addresses. They could be copycats of legitimate websites. Use a safe search tool such as McAfee SiteAdvisor to steer clear of risky sites.
8. **Keep up to date.** Keep all your software updated so you have the latest security patches. Turn on automatic updates so you don’t have to think about it, and make sure that your security software is set to run regular scans.

9. **Lookout for the latest scams.** Online threats are evolving all the time, so make sure you know what to look out for. Currently, “ransomware” is on the rise. This is when a hacker threatens to lock you out of all of your files unless you agree to pay a ransom. Stay on top of this and other threats by staying informed.
Appendix C: Materials Created for Ethnicity Manipulation

Mohamad Salim

CAREER OBJECTIVE
Insurance sales manager with 9+ years of experience driving profitability through strategic growth, leading teams, and quality control. Fiercely competitive in my approach to acquire business, and able to handle complex situations from a strategic and tactical perspective.

EXPERIENCE

Al-Salam Real Estate,
Sales Manager - Doha, Qatar / September 2014 - Present
- Supervise and manage a sales staff of 7; communicate job expectations and provide direction, support, and motivation to sales team to meet agreed targets and KPIs
- Enforce policies and procedures to ensure that my sales team achieves the customer service levels set by the organization, resulting in an annual increase of customer satisfaction by 5% per year
- Improve and maintain operational and profit objectives exceeding $13,000,000 in sales monthly within the district; implement company business plan and provide information for future improvement to the business manager

Mecca Insurance Company,
Sales Assistant - Jeddah, Saudi Arabia / June 2009 – August 2014
- Analyzed the specific needs of customers via careful surveys to develop data driven pitches, increasing profits by 15% over two years
- Trained two new employees in industry best practices for strategic pitching, presenting, and deal closing; new employee performance was 6% higher than average
- Fostered strong relationships with customers and strategic associates to elevate brand awareness through community initiatives such as Meals on Wheels
- Awarded “Outstanding Achievement” trophy for consistently performing above average sales in a company spanning 20 states

CONTACT
Email: Mohamadsalim20184@gmail.com
Phone: +1 (895) 555 5555
Address: 4397 Aaron Smith Drive
Hamburg, PA 17101
LinkedIn: linkedin.com/in/Mohamadsalim

SOFT SKILLS
Adaptability
Collaboration
Strong Work Ethic
Problem Solving
Team Building

HARD SKILLS
MS Excel Proficient
Business Management
Client Management
Fluent Arabic / English

EDUCATION
King Khalid University,
B.S. Business Management
Abha, Saudi Arabia
2009

King Khalid University,
A.A. Business Management
Abha, Saudi Arabia
2009
John Smith

CAREER OBJECTIVE

Insurance sales manager with 9+ years of experience driving profitability through strategic growth, leading teams, and quality control. Fiercely competitive in my approach to acquire business, and able to handle complex situations from a strategic and tactical perspective.

EXPERIENCE

Long & Thompson Real Estate,
Sales Manager - Concord, NH / September 2014 - Present
- Supervise and manage a sales staff of 7; communicate job expectations and provide direction, support, and motivation to sales team to meet agreed targets and KPIs
- Enforce policies and procedures to ensure that my sales team achieves the customer service levels set by the organization, resulting in an annual increase of customer satisfaction by 5% per year
- Improve and maintain operational and profit objectives exceeding $13,000,000 in sales monthly within the district; implement company business plan and provide information for future improvement to the business manager

TRP Insurance Company
Sales Assistant - Concord, NH / June 2009 – August 2014
- Analyzed the specific needs of customers via careful surveys to develop data driven pitches, increasing profits by 15% over two years
- Trained two new employees in industry best practices for strategic pitching, presenting, and deal closing; new employee performance was 6% higher than average
- Fostered strong relationships with customers and strategic associates to elevate brand awareness through community initiatives such as Meals on Wheels
- Awarded “Outstanding Achievement” trophy for consistently performing above average sales in a company spanning 20 states

CONTACT

Email: johnsmith20194@gmail.com
Phone: +1 (896) 555 5555
Address: 4397 Aaron Smith Drive Harrisburg, PA 17101
LinkedIn: linkedin.com/in/Johnsmith

SOFT SKILLS

Adaptability
Collaboration
Strong Work Ethic
Problem Solving
Team Building

HARD SKILLS

MS Excel Proficient
Business Management
Client Management
Fluent Spanish / English

EDUCATION

Southern New Hampshire University,
B.S. Business Management Manchester, NH 2009

Southern New Hampshire University,
A.A. Business Management Manchester, NH 2009
Appendix D: Ancillary Analyses Explored

Table 4

*Study 1 Means and Standard Deviations Arab and White Applicants’ Interview ratings at All Levels of Interview Competence.*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>White – Low Interview</td>
<td>54</td>
<td>3.06</td>
<td>.82</td>
</tr>
<tr>
<td>White – Med Interview</td>
<td>51</td>
<td>3.22</td>
<td>.71</td>
</tr>
<tr>
<td>White – High Interview</td>
<td>55</td>
<td>3.31</td>
<td>.90</td>
</tr>
<tr>
<td>Arab – Low Interview</td>
<td>55</td>
<td>2.76</td>
<td>.82</td>
</tr>
<tr>
<td>Arab – Med Interview</td>
<td>50</td>
<td>3.17</td>
<td>.79</td>
</tr>
<tr>
<td>Arab – High Interview</td>
<td>54</td>
<td>3.30</td>
<td>.91</td>
</tr>
</tbody>
</table>

*Figure 7. Warmth results for interview competence conditions by ethnicity. Error bars represent standard error of the mean. Means and Standard deviations below.*

Table 5
Study 1 Means and Standard Deviations Arab and White Applicants’ Warmth ratings at All Levels of Interview Competence.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>White – Low Interview</td>
<td>54</td>
<td>3.4</td>
<td>.96</td>
</tr>
<tr>
<td>White – Med Interview</td>
<td>51</td>
<td>3.3</td>
<td>.87</td>
</tr>
<tr>
<td>White – High Interview</td>
<td>55</td>
<td>3.5</td>
<td>.96</td>
</tr>
<tr>
<td>Arab – Low Interview</td>
<td>55</td>
<td>3.5</td>
<td>1.07</td>
</tr>
<tr>
<td>Arab – Med Interview</td>
<td>50</td>
<td>3.9</td>
<td>.93</td>
</tr>
<tr>
<td>Arab – High Interview</td>
<td>53</td>
<td>3.5</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Figure 8. Competence results for interview competence conditions by ethnicity. Error bars represent standard error of the mean.

Table 6

Study 1 Means and Standard Deviations Arab and White Applicants’ Competence Ratings at All Levels of Interview Competence.
### Table 7

*Study 2 Means and Standard Deviations Arab and White Applicants’ Interview Ratings at All Levels of Interview Competence and Diversity Training conditions.*

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>White – Low Interview</td>
<td>54</td>
<td>3.6</td>
<td>.79</td>
</tr>
<tr>
<td>White – Med Interview</td>
<td>51</td>
<td>3.6</td>
<td>.88</td>
</tr>
<tr>
<td>White – High Interview</td>
<td>55</td>
<td>3.6</td>
<td>.85</td>
</tr>
<tr>
<td>Arab – Low Interview</td>
<td>55</td>
<td>3.4</td>
<td>1.12</td>
</tr>
<tr>
<td>Arab – Med Interview</td>
<td>50</td>
<td>3.6</td>
<td>.90</td>
</tr>
<tr>
<td>Arab – High Interview</td>
<td>54</td>
<td>3.8</td>
<td>.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Training; White – Low Interview</td>
<td>21</td>
<td>3.26</td>
<td>3.26</td>
</tr>
<tr>
<td>Control Training; White – Med Interview</td>
<td>25</td>
<td>3.16</td>
<td>3.16</td>
</tr>
<tr>
<td>Control Training; White – High Interview</td>
<td>25</td>
<td>2.91</td>
<td>2.91</td>
</tr>
<tr>
<td>Control Training; Arab – Low Interview</td>
<td>22</td>
<td>3.47</td>
<td>3.47</td>
</tr>
<tr>
<td>Control Training; Arab – Med Interview</td>
<td>29</td>
<td>2.84</td>
<td>2.84</td>
</tr>
<tr>
<td>Control Training; Arab – High Interview</td>
<td>23</td>
<td>3.55</td>
<td>3.55</td>
</tr>
<tr>
<td>Perspective Taking; White – Low Interview</td>
<td>24</td>
<td>3.32</td>
<td>3.32</td>
</tr>
<tr>
<td>Perspective Taking; White – Med Interview</td>
<td>23</td>
<td>3.09</td>
<td>3.09</td>
</tr>
<tr>
<td>Perspective Taking; White – High Interview</td>
<td>24</td>
<td>3.56</td>
<td>3.56</td>
</tr>
<tr>
<td>Perspective Taking; Arab – Low Interview</td>
<td>19</td>
<td>2.75</td>
<td>2.75</td>
</tr>
<tr>
<td>Perspective Taking; Arab – Med Interview</td>
<td>23</td>
<td>2.71</td>
<td>2.71</td>
</tr>
<tr>
<td>Perspective Taking; Arab – High Interview</td>
<td>25</td>
<td>3.03</td>
<td>3.03</td>
</tr>
<tr>
<td>Goal Setting; White – Low Interview</td>
<td>20</td>
<td>3.13</td>
<td>3.13</td>
</tr>
<tr>
<td>Goal Setting; White – Med Interview</td>
<td>20</td>
<td>3.05</td>
<td>3.05</td>
</tr>
<tr>
<td>Goal Setting; White – High Interview</td>
<td>24</td>
<td>3.24</td>
<td>3.24</td>
</tr>
<tr>
<td>Goal Setting; Arab – Low Interview</td>
<td>23</td>
<td>3.38</td>
<td>3.38</td>
</tr>
<tr>
<td>Goal Setting; Arab – Med Interview</td>
<td>23</td>
<td>3.62</td>
<td>3.62</td>
</tr>
<tr>
<td>Goal Setting; Arab – High Interview</td>
<td>25</td>
<td>3.33</td>
<td>3.33</td>
</tr>
</tbody>
</table>
Hypothesis 1 Re-test at Each Interview Question Level (Study 1 Data).

**Question 1 Results.** To re-test Hypothesis 1 with only question 1 rather than averaged score, the 2x3 ANOVA on interview scores showed a significant main effect for the interview competence level manipulation, $F(2, 313) = 4.05, p < .05, \eta^2_p = .03$, suggesting that raters were successfully able to distinguishing between performance conditions (low, moderate, high). However, the main effect of ethnicity $F(1, 313) = 1.62, p = .20, \eta^2_p = .01$, and subsequent interaction effect $F(2, 313) = 2.19, p = .11, \eta^2_p = .01$, were both not significant, suggesting that interview ratings did not significantly differ by ethnicity, nor did ethnicity and interview performance group interact to influence Question 1 interview ratings.

**Question 2 Results.** To re-test Hypothesis 1 with only question 2 rather than averaged score, the 2x3 ANOVA on interview scores showed a significant main effect for the interview competence level manipulation, $F(2, 313) = 4.14, p < .05, \eta^2_p = .03$, suggesting that raters were successfully able to distinguishing between performance conditions (low, moderate, high). However, the main effect of ethnicity $F(1, 313) = 0.97, p = .32, \eta^2_p = .003$, and subsequent interaction effect $F(2, 313) = 1.86, p = .16, \eta^2_p = .01$, were both not significant, suggesting that interview ratings did not significantly differ by ethnicity, nor did ethnicity and interview performance group interact to influence Question 2 interview ratings.

**Question 3 Results.** To re-test Hypothesis 1 with only question 3 rather than averaged score, the 2x3 ANOVA on interview scores did not show a significant main effect for the interview competence level manipulation, $F(2, 313) = 2.44, p = .09, \eta^2_p = .02$. Also, the main effect of ethnicity $F(1, 313) = 0.18, p = .68, \eta^2_p = .001$, and subsequent interaction effect $F(2, 313) = 0.04, p = .96, \eta^2_p = .000$, were both not significant, suggesting that interview ratings did
not significantly differ by ethnicity, nor did ethnicity and interview performance group interact to influence Question 1 interview ratings.