PERSONNEL PSYCHOLOGY
2010, 63, 325–359

ARE HIGHLY STRUCTURED JOB INTERVIEWS RESISTANT TO DEMOGRAPHIC SIMILARITY EFFECTS?

JULIE M. McCARTHY
The University of Toronto

CHAD H. VAN IDDEKINGE
The Florida State University

MICHAEL A. CAMPION
Purdue University

This study examines the extent to which highly structured job interviews are resistant to demographic similarity effects. The sample comprised nearly 20,000 applicants for a managerial-level position in a large organization. Findings were unequivocal: Main effects of applicant gender and race were not associated with interviewers' ratings of applicant performance nor was applicant–interviewer similarity with regard to gender and race. These findings address past inconsistencies in research on demographic similarity effects in employment interviews and demonstrate the value of using highly structured interviews to minimize the potential influence of applicant demographic characteristics on selection decisions.

Employment interviews are one of the most common selection devices used by organizations. When structured techniques are employed, interviews are able to obtain impressive levels of predictive validity (e.g., Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Wiesner & Cronshaw, 1988). Nevertheless, there exists a seemingly persistent belief among academics, practitioners, and the general public that group-level characteristics, such as race and gender, can have an undue influence on selection decisions, such as job interview scores (Landy, 2008). Indeed, episodes of racism, sexism, and other forms of workplace discrimination are a common topic in the popular press (e.g., Cardona, 2009; Miley & Wheaton, 2009; Pear, 2009). For example, selection process discrimination claims with respect to race and gender have reached

We thank Phil Roth for his helpful comments and suggestions on an earlier version of this article. The authors gratefully acknowledge Mathilda du Toit and Yuk Fai Cheong for their assistance in DOS programming and model estimation. In addition, special thanks to Mathew H. Reider for his help with the data and John P. Trougakos for his help with HLM.

Correspondence and requests for reprints should be addressed to Julie McCarthy, The University of Toronto, Management, 1265 Military Trail, Toronto, Ontario, M1C1A4 Canada; mccarthy@utsc.utoronto.ca.

an all-time high (Equal Employment Opportunity Commission [EEOC], 2009).

The potential influence of demographic characteristics is particularly relevant for selection/promotion systems that incorporate employment interviews given the interpersonal nature of the interview situation. For instance, interviewees' performance—and interviewers' evaluations of that performance—could be influenced not only by the interviewee's demographics (i.e., a main effect) but also by the match (or mismatch) between the interviewee's demographics and the interviewer(s)'s demographics (i.e., an interaction effect). This latter situation is referred to *demographic similarity*, and relevant theories predict that people will evaluate others who have similar group-level characteristics (e.g., gender) to themselves more favorably than those who are less similar (Tsui, Egan, & O'Reilly, 1992). The potential for demographic similarity effects to occur is a serious concern, as they may result in unfavorable selection decisions for dissimilar applicants and act to increase the potential for litigation (Offerman & Gowing, 1993; Williamson, Campion, Malos, Roehling, & Campion, 1997). Demographic similarity effects may also cause interviewers from dissimilar groups to treat applicants differently, resulting in negative applicant reactions. These negative reactions, in turn, can have a variety of deleterious effects, including reduced test-taking motivation and lower job acceptance rates (Ryan, 2001; Saks & McCarthy, 2006). Finally, demographic similarity effects may reduce the predictive validity of the interview process by unduly influencing interview scores and subsequently reducing the impact of candidate knowledge, skills, abilities, and other characteristics (KSAOs; McFarland, Ryan, Sacco, & Kriska, 2004).

A considerable amount of research has examined the *main effects* of applicant demographic characteristics on the ratings they receive in interviews. Meta-analyses have revealed relatively small main effects with respect to applicant race (Huffcutt & Roth, 1998) and gender (Olian, Schwab, & Haberfeld, 1988), particularly when structured interview formats are employed (Huffcutt & Roth, 1998). Meta-analytic estimates with respect to applicant gender also yield small main effects (Olian et al., 1988). Although these studies are informative, they fail to consider the fact that interviews involve interactions between an applicant and one or more interviewers. Thus, it is possible that the demographic similarity between the applicant and the interviewer could impact subsequent interview scores. For this reason, recent interest in demographics and interviews has shifted away from simple main effects toward more sophisticated demographic similarity models (e.g., Buckley, Jackson, Bolino, Veres, & Feild, 2007; Goldberg, 2005; Sacco, Scheu, Ryan, & Schmitt, 2003).

Corresponding findings from this relatively new body of research have yielded mixed results.

We suggest that this inconsistent pattern of findings is due, in part, to the fact that prior research has examined a wide range of interview procedures, which vary considerably in their degree of standardization in terms of interview development, administration, and/or scoring. Much of the past research has also studied small samples of participants completing simulated interviews. Our study was designed to address these critical gaps in past research. We draw from theories of individuating information (i.e., Fiske & Neuberg, 1990; Kunda & Spencer, 2003; Kunda & Thagard, 1996) to propose that properly conducted interviews, which follow the key components of interview structure (Campion, Palmer, & Campion, 1997), will be resistant to the influence of applicant gender and race. We examine this issue using data from nearly 20,000 job applicants who underwent highly structured interviews.

Demographic Similarity Theory

Demographic similarity theory is concerned with the extent to which people use demographic variables, such as gender and race, to determine how similar they are to others (Tsui et al., 1992; Tsui & O'Reilly, 1989). Two interrelated theoretical perspectives form the basis of demographic similarity theory: the similarity-attraction paradigm (Byrne, 1961) and social identity theory (Ashforth & Mael, 1989; Tajfel & Turner, 1986). The similarity-attraction paradigm suggests that individuals regard others more positively when they are viewed as more similar to themselves because it is assumed that individuals with similar demographics will also have similar underlying attributes (Milliken & Martins, 1996). The social identity paradigm (Ashforth & Mael, 1989) suggests that our self-concepts originate from the groups, or social categories, to which we belong (e.g., demographic groups, occupational groups, sports groups). We determine our social identities by classifying ourselves into various groups, and we tend to identify with the groups that enable us to maintain positive self-identities. Inclusion of oneself in a particular category leads to more positive evaluations of in-group than of out-group members. These theories are based on the idea that "birds of a feather flock together" and predict that people will evaluate group members with similar demographic backgrounds (i.e., gender, race) more favorably. Applied to an interview context, these theories predict that demographic similarity between applicants and interviewers will lead to higher levels of interpersonal attraction and, in turn, more favorable outcomes for "similar" applicants.

There is considerable evidence that demographic similarity can influence work outcomes (Riordan, 2000). For example, demographic similarity has been found to lead to more positive employee relations and communication patterns and higher job satisfaction (Ensher & Murphy, 1997; Green, Anderson, & Shivers, 1996; Tsui & O'Reilly, 1989; Wesolowski & Mossholder, 1997). Findings are somewhat less consistent in evaluation contexts. For example, demographic similarity has been found to have no effect on performance ratings (Rotundo & Sackett, 1999; Waldman & Avolio, 1991), small effects on performance ratings (Pulakos, White, Oppler, & Borman, 1989), and moderate effects on performance ratings (McKay & McDaniel, 2006; Roth, Huffcutt, & Bobko, 2003).

Prior Research on Demographic Similarity Effects in Interviews

A number of researchers have examined the extent to which demographic similarity influences interview scores. We summarize this research in Table 1. For each study, we indicate the type of similarity examined, the study context and sample, the type of interview(s), and the key findings. The magnitude of observed effects are interpreted in a manner consistent with Cohen (1988). As shown, the findings of these studies are varied, with some reporting no effects (e.g., Graves & Powell, 1995; Sacco et al., 2003; Simas & McCarrey, 1979) and others reporting small to moderate effects (e.g., Buckley et al., 2007; Lin, Dobbins, & Farh, 1992; McFarland et al., 2004).

This variability may be due, in part, to the fact that a majority of studies have focused on simulated interviews (e.g., Buckley et al., 2007; Gallois, Callan, & Palmer, 1992; Simas & McCarrey, 1979) and assessment center ratings that have not teased apart the effects of interviews from other exercises (e.g., Fiedler, 2001; Walsh, Weinberg, & Fairfield, 1987). Although lab-based research possesses several advantages (e.g., increased control; Mook, 1983), simulated interviews may not capture the motivations and consequences that affect the conduct and evaluation of real interviews. Moreover, the findings of assessment center studies do not speak directly to demographic similarity effects in interviews. For example, although interviews and assessment center exercises share common elements, they also vary in the extent to which applicants must interact with interviewers/assessors and in what applicants are required to do (e.g., respond to interview questions vs. interact with other applicants in a leaderless group discussion).

The between-study differences in past interview similarity research may also be due to sampling error. Indeed, many studies have examined demographic similarity using modest samples of interviewees and/or interviewers (see Table 1). Small samples can give rise to sampling

TABLE 1
Prior Research on Demographic Similarity Effects in Job Interviews

Study	Type of similarity	Context and sample	Interview type	Key findings
Buckley, Jackson, Bolino, Veres, and Field (2007)	RACE -Black -White	LAB STUDY -20 assessors viewed videotapes of 73 officers applying for a real promotion	SITUATIONAL	RACE SIMILARITY -significant effect, small in magnitude RACE SIMILARITY × PANEL COMPOSITION -significant effect, small in magnitude
Elliott (1981)	GENDER	FIELD STUDY -400 applicants for selection at Irish banks -38 interviewers	UNSTRUCTURED	GENDER SIMILARITY -no significant effect
Fiedler (2001)	RACE -Hispanic -Non-Hispanic	FIELD STUDY -341 applicants for a sales position -total number of assessors not reported	UNSPECIFIED: -assessment center ratings included interviews; interview data not reported separately	RACE SIMILARITY -no significant effect
Gallois, Callan, and Palmer (1992)	GENDER	LAB STUDY -56 personnel officers viewed 6 videotapes of simulated interviews	UNSPECIFIED	GENDER SIMILARITY -no significant effect

continued

TABLE 1 (continued)

Study	Type of similarity	Context and sample	Interview type	Key findings
Goldberg (2005)	GENDER RACE -Black -White	COLLEGE RECRUITING INTERVIEWS -273 students applying for various jobs & companies through career services -45 interviewers	UNSPECIFIED: -interviews varied by different recruiters -type of interview varied and was not controlled	GENDER SIMILARITY -no significant effect RACE SIMILARITY -significant effect, small in magnitude
Graves and Powell (1995)	GENDER	COLLEGE RECRUITING INTERVIEWS -476 students applying for various jobs & companies through career services -483 interviewers	UNSPECIFIED: -interviews varied by different recruiters -type of interview varied and was not controlled	GENDER SIMILARITY -no significant effect
Graves and Powell (1996)	GENDER	COLLEGE RECRUITING INTERVIEWS -680 students applying for a various jobs & companies through career services -237 interviewers	UNSPECIFIED: -interviews varied by different recruiters -type of interview varied and was not controlled	GENDER SIMILARITY -no significant effect
Lin, Dobbins, and Farh (1992)	RACE -Black -White -Hispanic	FIELD STUDY -2,805 applicants for a custodial job -total number of interviewers not reported	PAST-BEHAVIORAL AND SITUATIONAL	RACE SIMILARITY -significant effect, small in magnitude

continued

TABLE 1 (continued)

Study	Type of similarity	Context and sample	Interview type	Key findings
McFarland, Ryan, Sacco, and Kriska (2004)	RACE -Black -White	FIELD STUDY -1,334 police officer applicants -21 interviewers	SITUATIONAL	RACE SIMILARITY -no significant effect RACE SIMILARITY × PANEL COMPOSITION -significant effect, small in magnitude
Prewett-Livingston, Field, Veres, and Lewis (1996)	RACE -Black -White	FIELD STUDY -153 police officers applying for promotion to rank of sergeant -24 interviewers	SITUATIONAL	RACE SIMILARITY -significant effect, medium in magnitude RACE SIMILARITY × PANEL COMPOSITION -significant effect, small in magnitude
Rand and Wexley (1975)	RACE -Black -White	LAB STUDY -160 undergraduate students viewed 2 simulated video taped interviews	UNSPECIFIED	RACE SIMILARITY -significant effect, medium in magnitude
Reid, Kleiman, and Travis (2001)	GENDER	LAB STUDY -180 undergraduate students read 6 simulated paper interview scenarios	UNSPECIFIED	GENDER SIMILARITY -no significant effects

continued

TABLE 1 (continued)

Study	Type of similarity	Context and sample	Interview type	Key findings
Sacco, Scheu, Ryan, and Schmitt (2003)	GENDER RACE -Black -White -Asian -Hispanic	COLLEGE RECRUITING INTERVIEWS -12,203 students applying for various jobs with a large manufacturing firm -708 interviewers	PAST-BEHAVIORAL	RACE SIMILARITY -no significant effect GENDER SIMILARITY -no significant effect
Simas and McCarrey (1979)	GENDER	LAB STUDY -28 individuals viewed 4 simulated videotaped interviews	UNSPECIFIED	GENDER SIMILARITY -no significant effect
Walsh, Weinberg, and Fairfield (1987)	GENDER	FIELD STUDY -1,035 candidates for a professional sales position in a large financial services organization -133 assessors	UNSPECIFIED: -assessment center included an interview but interview data not reported separately	GENDER SIMILARITY -significant effect, small in magnitude
Wiley and Eskilson (1985)	GENDER	LAB STUDY -109 undergraduate students read 2 simulated paper interview scenarios	UNSTRUCTURED	GENDER SIMILARITY -no significant effect

Note. We used Cohen's (1988) criteria to describe the effect sizes found in each study as reflecting small, medium, and large effects. (d: small = .20 – .49, medium = .50 – .79, large = >.80; η^2: small = .01 – .05; medium = .06 – .13; large = >.14).

error, which can produce observed effects that are substantially different from actual population effects (Schmidt & Hunter, 1996). In support of this possibility, studies that have examined relatively small numbers of interviewees and/or interviewers have tended to yield inconsistent results with respect to similarity effects. In contrast, the few large-sample studies that have been conducted (e.g., Lin et al., 1992; McFarland et al., 2004; Sacco et al., 2003) have found consistently small similarity effects.

Finally, the variation in past research may be a result of differences in interview structure, which we believe is an important factor with respect to demographic similarity effects. Some studies have not reported the amount of interview structure, and others have been unable to control the amount of structure due to the fact that the interviews were conducted by recruiters from different units or organizations (e.g., Fiedler, 2001; Gallois et al., 1992; Goldberg, 2005). In cases where structured interviews have been used, many of the key components were not followed. In their review paper, Campion et al. (1997) identified 15 key elements of interview structure. They maintain that properly designed structured interviews should contain the following components: (1) job analysis, (2) same questions, (3) limited prompting, (4) better questions, (5) longer interviews, (6) control of ancillary information, (7) limited questions from candidates, (8) multiple rating scales, (9) anchored rating scales, (10) detailed notes, (11) multiple interviewers, (12) consistent interviewers, (13) no discussion between interviews, (14) training, and (15) statistical prediction.

We found five demographic similarity studies that examined interviews that appeared to incorporate the majority of these elements. Lin et al. (1992) examined demographic effects among 2,805 Black, White, and Hispanic individuals applying for a custodial job. Each applicant was rated by a racially mixed panel of two interviewers (the total number of interviewers was not reported). Although demographic similarity effects were statistically significant across both the situational and past-behavioral interviews, the magnitude of the observed effects was small. This study provided a valuable starting point for research examining demographic similarity effects in highly structured interviews. However, not all of the key components of interview structure were followed. For example, the past-behavioral interview format did not use anchored rating scales, and the length of the situational interview was not reported. Moreover, as noted by the authors, the strong underrepresentation of White applicants (approximately 5%) resulted in low statistical power.

Prewett-Livingston, Feild, Veres, and Lewis (1996) examined 153 police officer candidates, each of whom was rated by a racially mixed panel of four interviewers (24 total interviewers). Findings indicated a significant similarity effect, which was medium in magnitude. Although the

researchers indicated use of a highly structured interview, they did not report factors such as the length of the interview and whether they controlled ancillary information, interviewer prompting, or questions from the candidate. Moreover, the study was based on a relatively small sample of primarily male police officers who were White or Black, which precluded an assessment of similarity effects with respect to gender and other racial minorities (e.g., Hispanics).

Sacco et al. (2003) examined 12,203 undergraduates who participated in real recruiting interviews. Students applied to a variety of jobs in a large manufacturing firm, and each student was given a one-on-one past-behavioral interview administered by one of 708 college recruiters. This study represented a particularly significant extension of past work, as it was based on a very large sample and hierarchical linear modeling (HLM) was used to analyze the data. This study also considered similarity effects with respect to both gender and to each of the four primary racial groups in the United States. (i.e., Asian, Black, Hispanic, and White). Interviews were highly structured but were conducted by a single interviewer rather than an interview panel. Findings revealed no evidence of racial or gender similarity effects. Given that the focus of this study was on students undergoing initial recruitment interviews for entry-level jobs, examination of whether these findings generalize to managerial-level employees and to other interview types would be valuable.

The fourth study was conducted by McFarland et al. (2004) and was based on 1,334 police officer candidates. Candidates underwent a situational interview administered by a racially mixed panel of three interviewers (21 total interviewers). A unique feature of this study was that it was longitudinal, and thus considered changes in interview ratings over time. In terms of results, interactions among applicant race, rater race, and the composition of the interview panel were statistically significant but small in magnitude. However, the data were analyzed using analysis of variance, which required computing an average overall rating for each applicant across the individual interviewers in each panel. Thus, the researchers were unable to examine demographic similarity between each applicant and each individual interviewer.

The most recent study in the area was conducted by Buckley et al. in 2007. In this study, 20 assessors evaluated videotapes showing 73 police officers responding to a single situational interview question. By employing a lab simulation, the racial composition of the interview panels was manipulated, such that all possible Black/White racial combinations were represented on the different interview panels. Racial similarity effects with respect to panel composition were found to be statistically significant, albeit small in magnitude. However, the extent to which this small sample of simulated interviews generalizes to face-to-face interviews in organizational contexts is uncertain. Moreover, the data were analyzed

using analysis of variance, which prevented the researchers from accounting for the nested structure of the data.

The Resistance of Highly Structured Interviews to Demographic Similarity Effects

As indicated, the nature and magnitude of demographic similarity effects on interview scores is not conclusive. Our goal was to conduct a robust test of this phenomenon by developing a set of highly structured interviews and assessing racial and gender similarity effects in a very large sample of real applicants. We draw on theories of individuating information to propose that properly designed, highly structured interviews will be resistant to demographic similarity effects. Specifically, we contend that demographic similarity effects are less likely to play a role because structured interviews increase the amount of individuating information available to, and used by, interviewers.

Three theories of individuating information have been advanced (Fiske & Neuberg, 1990; Kunda & Spencer, 2003; Kunda & Thagard, 1996). These theories share several key assumptions. First, they hold that when an individual meets a new person, cognitive processing begins, and the initial categorization of the individual is often based on group-level characteristics, such as race or gender (see review by Fiske, 1998). This categorization can cause perceivers to think, feel, and behave in a specific way toward the target (Fiske, Lin, & Neuberg, 1999). In particular, demographic similarity theory suggests that when the demographic characteristics of the individual are categorized as similar to oneself, more positive perceptions and evaluations are likely to ensue (Byrne, 1961; Tajfel & Turner, 1986; Tsui & O'Reilly, 1989).

Also common among these theories is the belief that impressions are based on more than just demographic information and can be influenced by individuating information. Applied to the workplace, individuating information is conceptualized as our knowledge about the job-related behaviors and attributes of a specific individual (Copus, 2005). It includes, but is not limited to, knowledge, skills, abilities, personality traits, and behaviors. Thus, when forming impressions of others, individuals integrate the full range of information known to characterize the individual, including demographic characteristics and individuating information (Fiske & Neuberg, 1990). Further, to the extent that individuating information about the person becomes available, is processed, and is used, this information can override initial perceptions when final judgments of the person are made. Research supports this proposition, as the more individuating information that becomes available, the less influence demographic characteristics tend to have (Kunda & Thagard, 1996).

We suggest that highly structured interviews facilitate the acquisition and use of individuating information, which, in turn, overrides initial perceptions and provides resistance against demographic similarity effects. This may be accomplished in at least three ways. First, to elicit individuating processes, the perceiver must be motivated to form an accurate impression of the target (Fiske & Neuberg, 1990; Kunda & Spencer, 2003). This motivation determines whether the perceiver stays with an initial category-based impression or whether he or she moves beyond group identity to focus on individuating information (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002). Empirical findings suggest that individuals do not tend to seek out individuating information on their own (Cameron & Trope, 2004) but rather try to conserve their energy and form an impression of a target as soon as they feel they have enough information to form a plausible evaluation (Epley & Gilovich, 2006). In other words, people tend to stop adjusting their initial impression too soon.

Several features of highly structured interviews are likely to motivate interviewers to form an accurate impression of candidates and to make it difficult for interviewers to stop adjusting too soon. Fiske and Neuberg (1990) noted that when perceivers expect that their judgments will be made known or compared to others' judgments, they are more motivated to present an accurate impression. In structured interviews, the use of panels increases interviewer motivation to attend to individuating information because interviewers must explain their ratings to others (Arvey & Campion, 1982; Tetlock & Boettger, 1989). In other words, the anticipation of discussion among raters should lead to the greater attention to individuating information. Further, all types of highly structured interviews make it difficult for interviewers to stop adjusting too soon because the interview is not complete until all relevant KSAOs have been assessed. In particular, interviews ask a series of predetermined, job-relevant questions, rate interviewees' responses using behaviorally anchored rating scales aligned with a particular question or dimension, and derive a final evaluation that reflects a statistical combination of ratings across questions/dimensions (Campion et al., 1997).

Second, theories of individuation assert that the more attention the perceiver pays to the target, the more likely it is that they will notice, remember, and use the information that is *inconsistent* with initial perceptions (Fiske & Neuberg, 1990; Kunda & Spencer, 2003). In fact, attention is conceptualized as a central mediator, such that motivation to obtain an accurate impression leads to increased attention to the target, which, in turn, facilitates the acquisition and use of individuating information (Fiske & Neuberg, 1990). This idea has been supported by research showing that when attentional resources are scarce, raters organize their impressions based on group-level stereotypes (Biesanz, Neuberg, Smith, Asher, & Judice, 2001; Gilbert & Hixon, 1991; Harris-Kern & Perkins, 1995). In

contrast, when evaluators are forced to attend to individuating information about the targets (e.g., retrieving and recording specific characteristics of the target), memory for the targets' stereotyped traits is inhibited (Dunn & Spellman, 2003).

Highly structured interviews are designed to focus interviewers' attention on the job-relevant content of interviewees' responses. Structured interviews also tend to take longer than less structured interviews and thus allow ample opportunity for interviewers to obtain the requisite individuating information (Campion et al., 1997). Interviewer note taking also helps to ensure that interviewers focus their attention on the target. Indeed, note taking has been found to reduce the impact of preexisting expectations (which, for example, may be influenced by demographic stereotypes) on interviewers' final evaluations of applicants (Biesanz, Neuberg, Judice, & Smith, 1999).

A third way in which individuating information can help to override potential demographic similarity effects is by ensuring that interviewers focus on information that is predictive of job performance. Kunda and Thagard (1996) highlighted the importance of the relevance of the individuating information to the judgment in question. Specifically, if the individuating information is relevant to the evaluation task, then it is more likely to be incorporated into evaluations of the target individual, thereby overriding the influence of group-level stereotypes. In support of this proposition, a number of studies have found that providing behaviorally relevant information about targets reduces the use of race and gender-based stereotypes in evaluations of current and future performance (Bodenhausen, Macrae, & Sherman, 1999; Fiske, 1998; Kunda & Thagard, 1996).

Within an interview context, it is important to ensure that raters focus on individuating information relevant to the job (Tetlock, Mitchell, & Murray, 2008). The questions in high-structure interviews are designed to measure KSAOs and behaviors identified from a job analysis. In addition, interviewers evaluate interviewees' responses against rating scales that describe low, moderate, and high descriptions or examples of each KSAO/behavior. These features, coupled with the fact that highly structured interviews attempt to minimize the extent to which applicants can express irrelevant information (e.g., limit the opportunity to ask questions during the interview), help interviewers obtain and evaluate relevant individuating information.

In sum, structured interviews possess several characteristics that would seem to enable interviewers to acquire and use individuating information, thereby forming impressions of applicants that are minimally affected by demographic similarity effects. It is important to note, however, that impressions of others are formed by simultaneously integrating initial category-based information (e.g., demographic similarity) and

individuating information (Kunda & Thagard, 1996). As such, demographic characteristics have the potential to influence interviewers' judgments at any stage of the interview process, regardless of how much individuating information has already been obtained (Kunda & Spencer, 2003; Kunda & Thagard, 1996; Wessel & Ryan, 2008). This can occur, for example, when the individuating information is ambiguous (Darley & Gross, 1983; Kunda & Sherman-Williams, 1993). This underscores the importance of ensuring that the entire interview process remains highly structured. The influence of demographic-based judgments can also change when different judgment tasks are used (Berndt & Heller, 1986; Jackson, Sullivan, & Hodge, 1993), such as the use of different interview formats (e.g., past-behavioral vs. situational). Thus, examining interviewer ratings with respect to different types of interviews is also important.

Current Study

Our goal was to conduct a robust test of the extent to which three widely used types of structured interviews (past-behavioral, situational, and experience-based) are resistant to demographic similarity effects. Based on the above discussion of theory and research on demographic similarity, individuation, and structured interviewing, we do not expect high-structure interviews that conform to the key components of structure to be subject to racial or gender similarity effects. In examining this general expectation, we address several critical gaps in past research to provide a more definitive test of demographic similarity effects.

First, we provide a theoretical foundation for the proposed lack of demographic similarity effects in job interviews. This is an important contribution because there has been an absence of strong theory in past work on similarity effects in structured interviews. Second, our data set includes a large and diverse sample of both applicants ($N = 19{,}931$) and interviewers ($N = 207$). This enabled us to fully explore the range of demographic similarity effects that may be present in real-world selection situations. Third, several researchers have highlighted the need for research on demographic similarity that focuses on managerial jobs (Lin et al., 1992; McFarland et al., 2004; Prewett-Livingston et al., 1996). To our knowledge, this is one of the first investigations to do so. Fourth, we examine three types of highly structured interviews, which may be of considerable value to organizations that may need to choose one or two of these interview formats to use in the selection process (Simola, Taggar, & Smith, 2007). Fifth, unlike past research (see Sacco et al., 2003 for an exception), we consider demographic similarity with respect to the four primary U.S. racial groups, as well as with respect to gender. Finally, our large sample of interviewees and interviewers allow us to use advanced

HLM techniques to assess main and interactive effects of gender and race. This approach allows for more accurate estimates of demographic similarity than what more traditional approaches provide (e.g., ANOVA, linear regression).

Method

Sample

Participants included 19,931 entry-level persons applying for professional positions with an agency of the U. S. government. These positions entail working with the public, government officials, and the business community. Selected employees would work in one of several different career tracks, including general management and specialty areas. Thirty-four percent of the sample was female, and 59% was male. The remaining 7% did not identify their gender. In terms of racial composition, 15,709 participants were White (79%), 1,437 were Asian (7%), 1,026 were Hispanic (5%), and 700 were Black (4%). The remaining 5% did not report their ethnicity. Participants were retained if they reported data on either their gender and/or race.

A total of 207 interviewers participated in this research. All interviews were conducted by a panel of two interviewers, who were randomly assigned to applicants. In terms of gender, 74 interviewers were women (37%), 115 were men (58%), and 5% did not identify their gender. In terms of ethnicity, 131 interviewers were White (63%), 34 were Black (19%), 8 were Asian (4%), and 6 (3%) were Hispanic. The remaining 11% of interviewers did not identify their ethnicity. As with the interviewees, interviewers were retained if they reported data on either their gender and/or race.

Structured Interviews

We took great care to ensure the interviews incorporated the 15 key components of interview structure (Campion et al., 1997). The first seven components of structure focus on the content of the interviews. We ensured that (1) the interviews were based on a comprehensive job analysis; (2) within the situational and past-behavioral interviews, the same questions were asked of each candidate, and within the experience-based interview, similar questions were asked of each candidate; (3) the use of prompts and follow-up questions was limited; (4) three different questioning techniques were employed (i.e., experienced-based, situational, past-behavioral); (5) each interview allowed sufficient time for interviewers to ask several questions; (6) ancillary information was controlled; and (7) candidates were encouraged to ask questions after the structured phase of the interview process was complete.

The remaining components of structure focus on the evaluation of interviewees' responses. We ensured that: (8) interviewers evaluated each dimension using behaviorally anchored rating scales; (9) descriptive scale anchors were derived from KSAO definitions, previously developed interviews and responses from previous candidates; (10) interviewers were trained on the importance of note taking during the interview process; (11) a panel of two interviewers evaluated each candidate; (12) the same set of interviewers conducted the interviews for each applicant; (13) the interviewers did not discuss candidates between interviews; (14) all interviewers were extensively trained to ensure proficiency in conducting and scoring the interview; and (15) statistical procedures (unit weighting) were used to combine ratings within each interview.

Experienced-based interview. The experience-based interview required applicants to answer questions about their qualifications, such as work experience and education (cf., Roth & Campion, 1992). The one difference between the experience-based interview and the other two interviews was that interviewers could choose questions from a predetermined set of questions rather than asking every candidate the exact same questions. Interviewers rated candidates on three questions that corresponded to the following three KSAOs: education and work experience, motivation to join the organization, and other relevant experience.

Situational interview. The situational interview required applicants to respond to hypothetical dilemmas that may be experienced on the job (cf., Latham, Saari, Pursell, & Campion, 1980). It contained nine questions that corresponded to the following nine KSAOs: planning and organizing, teamwork, adaptability, leadership, judgment, integrity, analytical skills, resourcefulness, and composure. Each question consisted of a base question as well as follow-up questions that challenged the candidate by eliminating obvious answers and/or by changing the situation. One interviewer asked the questions, but both interviewers took notes and then rated the candidate's answers at the end of the interview.

Past-behavioral interview. The past-behavioral interview required applicants to describe their behavior in past situations relevant to the job (cf., Janz, 1982; Pulakos & Schmitt, 1995). This interview contained eight questions that corresponded to the following eight KSAOs: planning and organizing, teamwork, adaptability, leadership, judgment, integrity, composure, and oral communication skills.

Procedure

We used a within-subjects design, whereby each candidate completed all three interviews. Due to legal and practical considerations associated with interviewing 19,931 individuals, all applicants were administered

the interviews in the same order: experienced-based interview, situational interview, and past-behavioral interview. Each interview lasted approximately 20 minutes, for a total of about 60 minutes. Each applicant was interviewed in person by a panel of two experienced human resource specialists. These raters received 2 full days of training lead by a group of consultants who held doctoral degrees in I-O psychology or human resources management. Training consisted of lecture, practice evaluations of videotaped candidates, and feedback. Each rater was also given a manual that included the assessment, training notes, and other work aids. In addition to frame-of-reference training, a portion of the lecture material covered rater errors, including leniency, severity, and central tendency.

Interviewers rated each dimension assessed within each interview on a unique 7-point scale. The dimension ratings from each interview were then averaged to create a total score. If interviewers disagreed on the total score by more than two points, they would discuss the candidate. In situations where interviewers discussed their ratings, they had the choice to retain or change their original ratings. The data suggest that many retained their original ratings. Thus, even in cases where raters discussed their ratings, there was still opportunity for between-rater variance in their final ratings. After the interviews were complete, demographic information for both job candidates and interviewers was obtained from organizational records.

Analytic Strategy

We used HLM with restricted maximum likelihood (RML) estimation to analyze the data. The use of HLM enabled us to control the nonindependence of the interview scores resulting from the fact that two interviewers rated each applicant and that each interviewer evaluated multiple applicants. Additional benefits of using HLM to examine nested (i.e., multilevel) data structures have been highlighted by several researchers (e.g., Bliese, 2002; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004), and the specific benefits of using HLM to analyze interview data have been outlined by Sacco et al. (2003). The dependent variable for the Level 1 unit of analysis was the mean ratings of each of the two interviewers who conducted each interview. There were 119,586 scores at this level (i.e., the 19,931 applicants received two scores, one from each interviewer, for each of the three interviews). These scores were cross-classified by two higher-order Level 2 units: applicant demographic characteristics (gender, race) and interviewer demographic characteristics (gender, race). Thus, a cross-classified random effects model was estimated, with applicant and interviewer demographics nested within interview scores. The specific models to be estimated were beyond the current Windows version of HLM

6.0. Therefore, all analyses were run using advanced DOS programming in HLM 6.0 (Raudenbush et al., 2004).

HLM allowed us to assess whether our Level 2 variables (i.e., applicant and interviewer demographics) impacted outcomes at Level 1 (i.e., interview scores). This is analogous to testing for main effects of gender and race on interview scores. Given our interest demographic similarity effects, we also examined the interactions between applicant and interviewer demographic characteristics. These interactions are meaningful when the data are not centered, because a dichotomous coding strategy was used for both gender and racial effects (Sacco et al., 2003). Thus, we assessed uncentered, dichotomous variables for all analyses.

To facilitate interpretation of the findings, we compared seven gender and race subgroups. For the analyses involving gender, each applicant and interviewer was coded as 1 (male) or a 0 (female). This enabled an assessment of (a) the main effect of applicant gender on interview scores, (b) the main effect of interviewer gender on interview scores, and (c) the interaction between applicant and interviewer gender on interview scores, which tested for demographic similarity effects. The remaining six subgroups reflected all possible racial combinations: White/Black, White/Asian, White/Hispanic, Black/Asian, Black/Hispanic, and Asian/Hispanic. Using gender as an example, the following equations were estimated:

$$\textit{Level 1}: \text{Rating} = \beta_{ojk} + r_{ijk}$$

$$\textit{Level 2}: \beta_{oj} = \gamma_{00} + \gamma_{01}(sex_{app}) + y_{10} + y_{11}(sex_{int}) + \gamma_{01}(sex_{app})(sex_{int}) + u_{oj} + u_{ok}.$$

The L1 equation predicts applicants' interview scores based on the mean interview score within each of the j applicants (β_{ojk}) and the within-cell random effects for interview scores (r_{ijk}). The L2 equation models the main effect of applicant sex [γ_{01} (sex_{app})], the main effect of interviewer sex [$y_{11}(sex_{int})$], and interaction effect of applicant and interviewer sex [$\gamma_{01}(sex_{app})(sex_{int})$] on interview scores, and includes the intercept (γ_{oo}) and the residual random effects of applicants' and interviewers' demographic characteristics (u_{oj} and u_{ok} respectively).

Results

Descriptive statistics for interview scores for each combination of applicant and interviewer race are shown in Table 2, and the statistics for each combination of applicant and interviewer gender are shown in Table 3. The internal consistency reliability for scores on the experience-based interview was .79, for the situational interview was .90, and for the past-behavioral interview was .86. The intraclass correlations (C,2) for

TABLE 2
Descriptive Statistics of Interview Ratings for Each Applicant–Interviewer Race Combination

	Interviewer race				
Interview type	White	Black	Asian	Hispanic	Overall
EXPERIENCE-BASED					
White applicant					
M	5.14	5.14	5.17	5.10	5.14
SD	.73	.71	.63	.69	.71
N	21,485	5,632	2,027	885	31,371
Black applicant					
M	5.31	5.57	5.37	5.35	5.39
SD	.75	.71	.63	.66	.73
N	713	273	67	19	1,119
Asian applicant					
M	5.23	5.33	5.26	5.17	5.25
SD	.68	.74	.62	.70	.68
N	1,588	388	140	42	2,244
Hispanic applicant					
M	5.19	5.33	5.32	5.40	5.24
SD	.69	.67	.52	.69	.67
N	1,045	297	101	35	1,544
Overall					
M	5.15	5.18	5.19	5.12	5.16
SD	.72	.71	.63	.69	.72
N	24,929	6,619	2,347	984	34,879
SITUATIONAL					
White applicant					
M	4.93	4.93	4.88	4.98	4.93
SD	.72	.69	.63	.69	.71
N	21,485	5,632	2,027	885	30,029
Black applicant					
M	4.97	5.21	4.99	5.23	5.05
SD	.75	.78	.66	.76	.76
N	713	273	67	19	1,072
Asian applicant					
M	4.85	4.95	4.81	5.03	4.87
SD	.73	.73	.76	.65	.73
N	1,588	388	140	42	2,158
Hispanic applicant					
M	4.87	4.94	4.94	5.22	4.90
SD	.73	.70	.55	.67	.72
N	1,045	297	101	35	1,478
Overall					
M	4.92	4.94	4.88	4.99	4.93
SD	.72	.70	.64	.69	.71
N	24,929	6,619	2,437	984	34,879

continued

TABLE 2 (continued)

Interview type	Interviewer race: White	Black	Asian	Hispanic	Overall
PAST-BEHAVIORAL					
White applicant					
M	5.15	5.16	5.13	5.16	5.15
SD	.66	.63	.58	.62	.65
N	21,485	5,632	2,027	885	30,029
Black applicant					
M	5.24	5.49	5.25	5.41	5.31
SD	.71	.64	.62	.58	.69
N	713	273	67	19	1,072
Asian applicant					
M	5.12	5.24	5.09	5.21	5.14
SD	.65	.67	.74	.59	.66
N	1,588	388	140	42	2,158
Hispanic applicant					
M	5.16	5.19	5.25	5.35	5.18
SD	.61	.64	.57	.65	.62
N	1,045	297	101	35	1,478
Overall					
M	5.15	5.18	5.13	5.18	5.15
SD	.66	.64	.59	.62	.65
N	24,929	6,619	2,437	984	34,879

the mean interview ratings of the two raters in each interview were also high for the experience-based (.79), situational (.80), and past-behavioral (.82) interviews.

Confirmatory Factor Analyses

Prior to examining the existence of demographic similarity effects across the three interview formats, it was important to determine whether the formats were indeed distinguishable empirically. To do so, we conducted confirmatory analyses using AMOS 16.0 (Arbuckle, 2005) to examine the underlying structure of the interview ratings. Maximum likelihood estimation procedures were used and three indices were employed to assess the fit of the models: the chi-square index, the standardized root mean residual (SRMR; Hu & Bentler, 1999), and the comparative fit index (CFI; Bentler, 1990). This combination of fit indices ensured the inclusion of an index that considers how much variance is explained in light of how many degrees of freedom are used (i.e., SRMR) as well as an index that is a direct function of how much variance is explained by the model (i.e.,

TABLE 3
Descriptive Statistics of Interview Ratings for Each Applicant–Interviewer Gender Combination

	Interviewer		
Interview type	Female	Male	Overall
EXPERIENCE-BASED			
Female applicant			
M	5.26	5.27	5.27
SD	.66	.68	.67
N	7,964	5,315	13,446
Male applicant			
M	5.13	5.08	5.10
SD	.74	.73	.73
N	8,996	14,297	23,293
Overall			
M	5.18	5.15	5.16
SD	.72	.71	.71
N	14,311	22,261	36,572
SITUATIONAL			
Female applicant			
M	5.00	5.01	5.00
SD	.69	.68	.69
N	5,315	7,964	13,279
Male applicant			
M	4.89	4.88	4.89
SD	.72	.72	.72
N	8,996	14,297	23,293
Overall			
M	4.93	4.93	4.93
SD	.71	.71	.71
N	14,311	22,261	36,572
PAST-BEHAVIORAL			
Female applicant			
M	5.27	5.25	5.26
SD	.61	.61	.61
N	5,315	7,964	13,279
Male applicant			
M	5.13	5.07	5.10
SD	.67	.65	.66
N	8,996	14,297	23,293
Overall			
M	5.19	5.14	5.16
SD	.65	.64	.65
N	14,311	22,261	36,572

CFI). In the case of the SRMR, values approaching .00 indicate a good fit. For the CFI, values approaching 1.0 indicate a good fit. The mean dimensions ratings (averaged across the interviewers) within each interview format served as the input for these analyses.

We first tested a model where interviewers' ratings were specified to load on one of three factors that corresponded to the three interview formats (i.e., experience-based, situational, and past-behavioral). Findings indicated that this model achieved an acceptable fit to the data ($\chi^2_{(167)} = 23{,}171.6, p < .01$; SRMR = .05, CFI = .90). We compared this to a model in which the questions from the three interviews were specified to load on a single factor. We created this factor by fixing the covariances among the three interview factors to 1.0, and thus this model is nested within the three-factor model. This unidimensional model tested the possibility that applicants performed similarly (and/or interviewers evaluated that performance similarly) across all three interview formats rather than exhibiting distinct performance on the three interviews. Findings indicated that the three-factor model provided a significantly better fit to the data ($\Delta\chi^2_{(3)} = 8{,}289.96, p < .001$) than did the single-factor model ($\chi^2_{(170)} = 46{,}011.1$, $p < .01$; SRMR = .25, CFI = .80). This provides support for examining the three interview formats separately.

Hierarchical Linear Modeling Results

Results from HLM analyses are presented in Tables 4–7. The first section of each table presents findings for the experience-based interviews, the second section for the situational interviews, and the third section for the past-behavioral interviews. In each case, main effects of applicant demographics on interview scores, interviewer demographics on interview scores, and the interaction between applicant and interview demographics on interview scores are reported.[1] We also computed effect size estimates, in the form of Cohen's *d* and pseudo R^2 values, following the multilevel modeling effect size computation provided by McNulty, O'Mara, and Karney (2008).[2] It is noteworthy that each of these analyses were conducted on a sample large enough to find even the minutest effects. Indeed, the statistical power to detect an effect size of less than .05 (or

[1]Analyses examining the effects of interview panel composition on interview scores were also conducted. In doing so, situations in which (a) one interviewer was the same race as the applicant, (b) both interviews were the same race as the applicant, and (c) neither interviewer was the same race as the applicant were considered. These analyses were also conducted with respect to interviewer gender. Findings were consistent with our previous results: panel composition did not have a meaningful impact on interview scores.

[2]The size of the effects (*r*) for each analysis was estimated using the formula provided by McNulty et al. (2008). These *r*s were then converted into estimates of Cohen's *d*.

TABLE 4
HLM Analyses of Race Similarity Effects (White vs. Black) on Interview Scores

	Experience-based interview				Situational interview				Past-behavioral interview			
	b	*SE*	*t*	*d*	*b*	*SE*	*t*	*d*	*b*	*SE*	*t*	*d*
Level 1: Intercept	5.19	.06	89.97***		4.92	.06	88.11***		5.19	.05	103.81***	
Level 2: Main effects												
Applicant race (AR)	−.03	.04	−.74	.01	−.01	.04	−.17	.00	−.05	.04	−1.27	.01
Interviewer race (IR)	−.03	.04	−.65	.01	.03	.04	.65	.01	.00	.04	−.04	.00
Level 2: Interaction AR × IR	.04	.03	1.17	.01	.00	.03	.00	.00	.02	.03	.93	.01
Variance estimates:												
Interview score (Level 1)	.10	.31			.09	.30			.07	.26		
Applicant (Level 2 row)	.38	.62			.38	.62			.33	.57		
Interviewer (Level 2 column)	.02	.13			.01	.12			.01	.11		
Pseudo R^2	.00				.00				.00			

Note. *N* (Level 1) = 28,774; *N* (Level 2 applicants) = 16,304; *N* (Level 2 interviewers) = 164. Race was coded as 1 = *White*, 2 = *Black*.
b = unstandardized beta coefficients; *SE* = standard error; *t* = t-ratio for each effect; *d* = effect size; Pseudo R^2 = the amount of variance in interview scores accounted for by applicant and interviewer gender main effects and interactions.
***$p < .001$.

TABLE 5
HLM Analyses of the Effects of Applicant and Interviewer Race (White vs. Asian) on Interview Scores

	Experience-based				Situational interview				Past-behavioral interview			
	b	*SE*	*t*	*d*	*b*	*SE*	*t*	*d*	*b*	*SE*	*t*	*d*
Level 1: Intercept	5.14	.07	72.13***		4.97	.07	70.53***		5.24	.06	84.33***	
Level 2: Main effects												
Applicant race (AR)	.00	.05	.10	.00	.03	.05	.60	.01	−.06	.04	−1.43	.02
Interviewer race (IR)	.05	.06	.87	.01	−.04	.06	−.60	.01	−.04	.05	−.81	.01
Level 2: Interaction AR × IR	−.04	.04	−.96	.01	−.04	.04	−1.29	.01	.02	.03	.50	.01
Variance estimates:												
Interview score (Level 1)	.09	.30			.09	.30			.07	.26		
Applicant (Level 2 row)	.37	.61			.37	.61			.31	.56		
Interviewer (Level 2 column)	.02	.13			.02	.13			.01	.11		
Pseudo R^2	.00				.00				.00			

Note. *N* (Level 1) = 27,142; *N* (Level 2 applicants) = 16,840; *N* (Level 2 interviewers) = 139. Race was coded as 1 = *White*, 2 = *Asian*.
b = unstandardized beta coefficients; *SE* = standard error; *t* = t-ratio for each effect; *d* = effect size; Pseudo R^2 = the amount of variance in interview scores accounted for by applicant and interviewer gender main effects and interaction effects.
$^{***}p < .001$.

TABLE 6
HLM Analyses of the Effects of Applicant and Interviewer Race (White vs. Hispanic) on Interview Scores

	Experience-based interview				Situational interview				Past-behavioral interview			
	b	*SE*	*t*	*d*	*b*	*SE*	*t*	*d*	*b*	*SE*	*t*	*d*
Level 1: Intercept	5.13	.10	49.81***		4.97	.10	51.99***		5.08	.09	59.06***	
Level 2: Main effects												
Applicant race (AR)	−.02	.07	−.31	.00	−.09	.07	−1.20	.01	.01	.06	.21	.00
Interviewer race (IR)	−.02	.09	−.20	.00	−.04	.09	−.51	.01	.07	.08	.88	.01
Level 2: Interaction AR × IR	.05	.07	.70	.01	.08	.06	1.26	.02	−.02	.06	−.40	.00
Variance estimates:												
Interview score (Level 1)	.09	.31			.09	.31			.06	.25		
Applicant (Level 2 row)	.37	.61			.38	.62			.32	.56		
Interviewer (Level 2 column)	.02	.15			.02	.13			.01	.12		
Pseudo R^2	.00				.00				.00			

Note. *N* (Level 1) = 23,019; *N* (Level 2 applicants) = 16,463; *N* (Level 2 interviewers) = 137. Race was coded as 1 = *White*, 2 = *Hispanic*. *b* = unstandardized beta coefficients; *SE* = standard error; *t* = t-ratio for each effect; *d* = effect size; Pseudo R^2 = the amount of variance in interview scores accounted for by applicant and interviewer gender main effects and interaction effects.

***$p < .001$.

TABLE 7
HLM Analyses of the Effects of Applicant and Interviewer Gender on Interview Scores

	Experience-based interview				Situational interview				Past-behavioral interview			
	b	*SE*	*t*	*d*	*b*	*SE*	*t*	*d*	*b*	*SE*	*t*	*d*
Level 1: Intercept	5.08	.04	138.24***		4.87	.04	135.35***		5.08	.03	160.16***	
Level 2: Main effects												
Applicant gender (AG)	.05	.02	3.03**	.03	.05	.02	3.37**	.03	.05	.01	3.66***	.04
Interviewer gender (IG)	.03	.02	1.24	.01	.01	.02	.31	.00	.02	.20	1.15	.01
Level 2: Interaction AG × IG	−.01	.01	−.57	.01	−.01	.01	−.79	.01	−.01	.01	−.88	.01
Variance estimates:												
Interview score (Level 1)	.10	.31			.09	.30			.07	.26		
Applicant (Level 2 row)	.38	.62			.39	.62			.32	.57		
Column (Level 2 column)	.02	.12			.01	.12			.01	.11		
Pseudo R^2	.00				.00				.00			

Note. *N* (Level 1) = 36,597; *N* (Level 2 applicants) = 18,541; *N* (Level 2 interviewers) = 192. Gender was coded as 1 = *male*, 2 = *female*. *b* = unstandardized beta coefficients; SE = standard error; *t* = t-ratio for each effect; *d* = effect size; Pseudo R^2 = the amount of variance in interview scores accounted for by applicant and interviewer gender main effects and interaction effects.

$^{**}p < .01$. $^{***}p < .001$.

an $r^2 < .01$) was .995, thus rendering the probability of a Type II error extremely low (Cohen, Cohen, West, & Aiken, 2003).

Tables 4–6 present the results for the racial analyses with the White/Black, White/Asian, and White/Hispanic subgroups. Findings were consistent across all groups—neither the main effects of applicant or interviewer demographics, nor the interaction between applicant and interviewer demographics, significantly influenced interview ratings. Not only were the effects nonsignificant, but the effect sizes were consistently below .10, rendering them extremely small (Cohen et al., 2003). Moreover, the pseudo R^2 values were zero, suggesting that none of the variance in interview scores could be attributed to the demographic effects. Analyses for the Black/Asian, Black/Hispanic, and Asian/Hispanic subgroups yielded an identical pattern of nonsignificant findings and are available from the first author upon request.

Table 7 presents the results for gender. The main effect of interviewer gender on applicant scores was nonsignificant across all three interviews, as was the interaction between applicant and interviewer gender. A significant main effect for applicant gender was found across all three interviews, such that females scored slightly higher than males. However, the magnitude of these effects was extremely small ($d = .03 - .04$). Moreover, R^2 values were zero across all interview types, indicating that demographic effects did not contribute to the variance in interview scores. The overall findings were unequivocal: demographic similarity effects did not have a meaningful impact on any of the interview scores.

Discussion

The extent to which demographic variables influence personnel decisions can have important consequences with respect to fairness, diversity, and legal defensibility (and perhaps even construct and/or criterion-related validity). We used theories that focus on individuating information as a basis to propose that when the key components of structure are adhered to, demographic similarity effects are unlikely to occur in employment interviews. We tested this proposition in a large sample of applicants for managerial positions. Demographic similarity was considered with respect to gender and all four primary racial groups in the U.S. Findings were robust and suggest that demographic similarity effects in highly structured interviews were trivial. This is an important finding because it suggests that, in addition to obtaining impressive levels of predictive validity (Huffcutt & Arthur, 1994; McDaniel et al., 1994; Wiesner & Cronshaw, 1988), structured interviews can minimize or eliminate potential bias with respect to demographic similarity between applicants and interviewers.

Implications for Theory and Practice

We drew from theories of individuating information to posit that highly structured interviews facilitate the acquisition and use of individuating information and are thereby resistant to the effects of demographic similarity. Results were unequivocal and provided strong support for this proposition. This provides a solid theoretical basis for the small similarity effects that have been found in past studies of structured interviews. Theories of individuating information may also extend to other personnel selection and human resource practices, such as letters of recommendation and performance appraisals. In particular, theories of individuating information assert that for individuating information to be obtained and used, raters should be motivated to form accurate impressions of the target, and raters should focus their attention on job-relevant behaviors (Fiske & Neuberg, 1990; Kunda & Spencer, 2003). In a performance appraisal context, rater motivation could be facilitated by increasing accountability through the use of multiple raters (i.e., 360° feedback). Rater attention to job-relevant behaviors could be facilitated by basing ratings on a set of predetermined job-relevant dimensions of behavior and by using behaviorally anchored scoring techniques. Similar techniques could be used to develop structured and standardized letters of recommendation that are based on multiple raters and evaluate candidates on job-relevant attributes.

Our study also highlights possible factors for why inconsistent results have been reported in past studies of demographic similarity in interviews. First, past studies have varied considerably on the amount of interview structure. Even in cases where highly structured interviews have been examined, there is no evidence of a study that has followed all of the key components of interview structure. It is therefore possible that structured interviews are only resistant to demographic similarity effects when all, or most, of the components of structure are followed. Second, past studies have varied widely on important factors such as study design (e.g., lab simulations vs. field studies). These methodological differences may explain, in part, the inconsistent findings of past work. A third possibility is that sampling error may have contributed to between-study because a majority of past studies were based on samples considerably smaller than that obtained in this study. Finally, the statistical techniques used to analyze past work may have contributed to between-study differences because prior studies have tended to rely on ANOVA-based techniques, which do not account for the nested nature of the datasets that are common in this area and may thus overestimate similarity effects (see Sacco et al., 2003). Combined, these factors highlight the importance of using large samples of job applicants and HLM analyses to derive the most accurate estimates of similarity effects.

From a practical perspective, our findings challenge the frequent assumption made by academics, practitioners, and the general public that demographic characteristics have a substantial impact on interview scores. Our results suggest that organizations that adopt carefully administered interviews that conform to the key components of structure can minimize concerns of applicant discrimination on the basis of gender and race. The use of highly structured interviews will also help to facilitate the selection of a diverse workforce, as well as act to reduce litigation concerns. Further, the racial composition of the panel did not affect interview scores. Thus, although the use of a diverse panel of raters may facilitate the attraction of diverse candidates (Avery & McKay, 2006), panel diversity (or lack thereof) is not associated with subsequent scores. Our findings also indicate that experience-based, situational, and past-behavioral interview formats are able to provide unique assessment information and yet are equally resistant to demographic similarity effects. Thus, the use of these highly structured interview formats, independently or in combination, can minimize the potential for demographic similarity effects to occur.

Limitations

Although this study was characterized by several notable strengths, it also contains certain limitations. Our goal was to examine whether highly structured interviews are resistant to demographic similarity effects, and this was the first study we know of to examine similarity effects across three commonly used structured interview formats. Due to practical considerations associated with interviewing 20,000 job candidates, the three interviews were administered to each applicant in the exact same order. This ordering was carefully planned to facilitate the logical flow of the interviews and also helped to ensure that every candidate was treated in exactly the same manner, which increases the legal defensibility of the interview process.

At the same time, the consistent ordering of the interviews does not preclude the possibility that the (lack of) demographic similarity effects in the experienced-based interview influenced the effects observed for the situational and past-behavioral interviews. However, as previously described, demographic information has the potential to influence interviewers' judgments at *any* stage of the interview process, regardless of how much individuating information has already been obtained (Kunda & Spencer, 2003; Kunda & Thagard, 1996; Wessel & Ryan, 2008). Thus, even if individuating information was obtained and used for the first interview, demographic characteristics still could have influenced subsequent interviews. Moreover, the CFA results indicate that the three interviews were empirically distinct, which suggests that interviewers considered

each interview separately rather than being influenced by some general impression (e.g., caused by a similarity effect) across all three interviews. Nevertheless, researchers interested in testing whether different interview formats are associated with different outcomes may wish to consider the use of counterbalanced designs.

Another potential limitation of the current work was that it was not possible to examine less structured interview formats. Unstructured interviews are less valid than their structured counterparts, and both ethical and legal concerns surround their use. Hence, many organizations, including the one that supported this study, do not use unstructured interviews for selection. Nevertheless, future research that directly compares similarity effects across interviews of varying structure may be advantageous, if it is possible to conduct such research in an actual selection context. For example, there may be certain boundary conditions with respect to interview structure that are necessary to avoid similarity effects. Length of the interview may be one such condition. Buckley et al. (2007) found some evidence of demographic similarity effects for simulated interviews that were based on a single-item and were therefore relatively short in length. Had the interview been longer, more individuating information would have been available and the effects may have diminished (Kunda & Spencer, 2003). Further, because interviews represent a selection method rather than construct (Arthur & Villado, 2008), it is possible that similarity effects may vary across interviews designed to measure different constructs.

Directions for Future Research

Given the rigor of interviews that conform to the essential components of structure, we anticipate that similar findings would be obtained if the influence of other potential group-level stereotypes (i.e., age, education, religion) were examined in highly structured interviews. Consistent with this proposition, there is growing evidence that age does not affect how applicants perform in structured interviews (e.g., Lin et al., 1992; Morgeson, Reider, Campion, & Bull, 2008). However, future research should examine the extent to which these findings generalize to broader attitudinal similarity variables, such as personality and values.

We also encourage future researchers to conduct more direct tests of the role of individuating information. In particular, it would be valuable to assess the relative impact of individuating information at different times in the interview process. Qualitative field studies that assess the underlying processes that operate with respect to individuating information in job interview contexts may also provide valuable insight. Moreover, research that examines whether the acquisition and use of individuating

information renders broader selection and human resource practices immune to demographic similarity effects would be valuable.

REFERENCES

Arbuckle JL. (2005). *AMOS 6.0 user's guide*. Chicago, IL: SPSS.

Arthur W, Villado AJ. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection. *Journal of Applied Psychology, 93*, 435–442.

Arvey RD, Campion JE. (1982). The employment interview: A summary and review of recent research. PERSONNEL PSYCHOLOGY, *35*, 281–322.

Ashforth BE, Mael F. (1989). Social identity theory and the organization. *Academy of Management Review, 14*, 20–39.

Avery DR, McKay PF. (2006). Target practice: An organizational impression management approach to attracting minority and female job applicants. PERSONNEL PSYCHOLOGY, *59*, 157–187.

Bentler PM. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Berndt TJ, Heller KA. (1986). Gender stereotypes and social inferences: A developmental study. *Journal of Personality and Social Psychology, 50*, 889–898.

Biesanz JC, Neuberg SL, Judice TN, Smith DM. (1999). When interviewers desire accurate impressions: The effects of notetaking on the influence of expectations. *Journal of Applied Social Psychology, 29*, 2529–2549.

Biesanz JC, Neuberg SL, Smith DM, Asher T, Judice TN. (2001). When accuracy-motivated perceivers fail: Limited attentional resources and the reemerging self-fulfilling prophecy. *Personality and Social Psychology* Bulletin, *27*, 621–629.

Bliese PD. (2002). Multilevel random coefficient modeling in organizational research: Examples using SAS and S-PLUS. In Drasgow F, Schmitt N (Eds.), *Measuring and analyzing behaviour in organizations: Advances in measurement and data analysis* (pp. 401–445). San Francisco: Jossey-Bass.

Bodenhausen GV, Macrae CN, Sherman JW. (1999). On the dialectics of discrimination: Dual processes in social stereotyping. In Chaiken S, Trope Y (Eds.), *Dual process theories in social psychology* (pp. 271–290). New York: Guilford.

Buckley MR, Jackson KA, Bolino MC, Veres JG, Feild HS. (2007). The influence of relational demography on panel interview ratings: A field experiment. PERSONNEL PSYCHOLOGY, *60*, 627–646.

Byrne D. (1961). Interpersonal attraction as a function of affiliation need and attitude similarity. *Human Relations, 14*, 283–289.

Cameron JA, Trope Y. (2004). Stereotype-biased search and processing of information about group members. *Social Cognition, 22,* 650–673.

Campion MA, Palmer DK, Campion JE. (1997). A review of structure in the selection interview. PERSONNEL PSYCHOLOGY, *50*, 655–702.

Cardona F. (2009). Denver firefighter sues department, claiming bias. *Denver Post*. January 7th, 2009.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen J, Cohen P, West S, Aiken L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd edition). Hillsdale, NJ: Erlbaum.

Copus D. (2005). A lawyer's view: Avoiding junk science. In Landy FJ (Ed.), *Employment discrimination litigation: Behavioral, quantitative and legal perspectives* (pp. 450–462). San Francisco: Jossey Bass.

Darley JM, Gross PH. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44*, 20–33.

Devine PG, Plant EA, Amodio DM, Harmon-Jones E, Vance SL. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82*, 835–848.

Dunn EW, Spellman BA. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology, 39*, 420–433.

Elliott AGP. (1981). Sex and decision making in the selection interview: A real-life study. *Journal of Occupational Psychology, 54*, 265–273.

Ensher EA, Murphy SE. (1997). Effects of race, gender, perceived similarity, and contact on mentor relationships. *Journal of Vocational Behavior, 50*, 460–481.

Epley N, Gilovich T. (2006). Are adjustments insufficient? *Personality and Social Psychology Bulletin, 30*, 447–460.

Equal Employment Opportunity Commission. (2009). *Charge statistics FY 1997 through FY 2007*. Retrieved on January 27, 2009, Available at http://www.eeoc.gov/stats/charges.html.

Fiedler AM. (2001). Adverse impact on Hispanic job applicants during assessment center evaluations. *Hispanic Journal of Behavioral Sciences, 23*, 102–110.

Fiske ST. (1998). Stereotyping, prejudice, and discrimination. In Gilbert DT, Fiske ST, Lindzey G (Eds.), *Handbook of social psychology* (4th edition, vol. 2, pp. 357–411). New York: McGraw-Hill.

Fiske ST, Lin M, Neuberg SL. (1999). The continuum model: Ten years later. In Chaiken S, Trope Y (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York: Guilford.

Fiske ST, Neuberg SL. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23*, 1–74.

Gallois C, Callan VJ, Palmer JM. (1992). The influence of applicant communication style and interviewer characteristics on hiring decisions. *Journal of Applied Social Psychology, 22*, 1041–1060.

Gilbert DT, Hixon JG. (1991). The trouble of thinking: Activation and application of stereotype beliefs. *Journal of Personality and Social Psychology, 60*, 509–517.

Goldberg CB. (2005). Relational demography and similarity-attraction in interview assessments and subsequent offer decisions: Are we missing something? *Group & Organization Management, 30*, 597–624.

Graves LM, Powell GN. (1995). The effect of sex similarity on recruiters' evaluations of actual applicants: A test of the similarity-attraction paradigm. PERSONNEL PSYCHOLOGY, *48*, 85–98.

Graves LM, Powell GN. (1996). Sex similarity, quality of the employment interview and recruiters' evaluation of actual applicants. *Journal of Occupational and Organizational Psychology, 69*, 243–261.

Green SG, Anderson SE, Shivers SL. (1996). Demographic and organizational influences on leader–member exchange and related work attitudes. *Organizational Behavior and Human Decision Processes, 66*, 203–214.

Harris-Kern MJ, Perkins P. (1995). Effects of distraction on interpersonal expectancy effects: A social interaction test of the cognitive busyness hypothesis. *Social Cognition, 13*, 163–182.

Hu L, Bentler PM. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling, 6*, 1–55.

Huffcutt AI, Arthur W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*, 184–190.

Huffcutt AI, Roth PL. (1998). Racial group differences in employment interview evaluations. *Journal of Applied Psychology, 83*, 179–189.

Jackson LA, Sullivan LA, Hodge CN. (1993). Stereotype effects of attributions, predictions, and evaluations: No two social judgments are quite alike. *Journal of Personality and Social Psychology, 65*, 69–84.

Janz T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67*, 577–580.

Kunda Z, Sherman-Williams B. (1993). Stereotypes and the construal of individuating information. *Personality and Social Psychology Bulletin, 19*, 90–99.

Kunda Z, Spencer SJ. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129*, 522–544.

Kunda Z, Thagard P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel constraint satisfaction theory. *Psychological Review, 103*, 284–308.

Landy FJ. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology, 1*, 379–392.

Latham GP, Saari LM, Pursell ED, Campion MA. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422–427.

Lin T, Dobbins GH, Farh J. (1992). A field study of race and age similarity effects on interview ratings in conventional and situational interviews. *Journal of Applied Psychology, 77*, 363–371.

McDaniel MA, Whetzel DL, Schmidt FL, Maurer SD. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*, 599–616.

McFarland LA, Ryan AM, Sacco JM, Kriska SD. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of Management, 30*, 435–452.

McKay PF, McDaniel MA. (2006). A reexamination of black-white mean differences in work performance: More data, more moderators. *Journal of Applied Psychology, 91*, 538–554.

McNulty JK, O'Mara EM, Karney BR. (2008). Benevolent cognitions as a strategy of relationship maintenance: "Don't sweat the small stuff" . . . . But it is not all small stuff. *Journal of Personality and Social Psychology, 94*, 631–646.

Miley M, Wheaton K. (2009). Agencies not only lack black workers, they pay them less; Mehri, NAACP say don't blame the victims, lay groundwork for suit. *Advertising Age, 80*, 1–23.

Milliken FJ, Martins LL. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *The Academy of Management Review, 21*, 402–433.

Mook DG. (1983). In defense of external invalidity. *American Psychologist, 38*, 379–387.

Morgeson FP, Reider MH, Campion MA, Bull RA. (2008). Review of research on age discrimination in the employment interview. *Journal of Business and Psychology, 22*, 223–232.

Offerman LR, Gowing MK. (1993). Personnel selection in the future: The impact of changing demographics and the nature of work. In Schmitt N, Borman WC (Eds.), *Personnel selection in organizations* (pp. 385–417). San Francisco: Jossey-Bass.

Olian JD, Schwab DP, Haberfeld Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of experimental studies. *Organizational Behavior and Human Decision Processes, 41*, 180–195.

Pear R. (2009, January 5). Justices' ruling in discrimination case may draw quick action by Obama. *New York Times*, A13.

Prewett-Livingston AJ, Feild HS, Veres III JG, Lewis PM. (1996). Effects of race on interview ratings in a situational panel interview. *Journal of Applied Psychology, 81*, 178–186.

Pulakos ED, Schmitt N. (1995). Experience-based and situational interview questions: Studies of validity. PERSONNEL PSYCHOLOGY, *48*, 289–308.

Pulakos ED, White LA, Oppler SH, Borman WC. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology, 74*, 770–780.

Rand TM, Wexley KN. (1975). Demonstration of the effect, "similar to me," in simulated employment interviews. *Psychological Reports, 36,* 535–544.

Raudenbush SW, Bryk AS, Cheong YF, Congdon R, du Toit M. (2004). *HLM 6.0: Hierarchical linear and non-linear modeling*. Lincolnwood, IL: Scientific Software International.

Reid PT, Kleiman LS, Travis CB. (2001). Attribution and sex differences in the employment interview. *The Journal of Social Psychology, 126*, 205–212.

Riordan CM. (2000). Relational demography within groups: Past developments, contradictions, and new directions. In Ferris GR (Ed.), *Research in personnel and human resources management* (vol. 19, pp. 131–173). Greenwich, CT: JAI Press.

Roth PL, Campion JE. (1992). An analysis of the predictive power of the panel interview and pre-employment tests. *Journal of Occupational and Organizational Psychology, 65*, 51–60.

Roth PL, Huffcutt AI, Bobko P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology, 88*, 694–706.

Rotundo M, Sackett PR. (1999). Effect of rater race on conclusions regarding differential prediction in cognitive ability tests. *Journal of Applied Psychology, 84*, 815–822.

Ryan AM. (2001). Explaining the black-white test score gap: The role of test perceptions. *Human Performance, 14*, 45–75.

Sacco JM, Scheu CR, Ryan AM, Schmitt N. (2003). An investigation of race and sex similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology, 88*, 852–865.

Saks AM, McCarthy JM. (2006). Effects of discriminatory interview questions and gender on applicant reactions. *Journal of Business and Psychology, 21*, 175–191.

Schmidt FL, Hunter JE. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*, 199–223.

Simas K, McCarrey MW. (1979). Impact of recruiter authoritarianism and applicant sex on evaluation and selection decisions in a recruitment interview analogue study. *Journal of Applied Psychology, 64*, 483–491.

Simola SK, Taggar S, Smith G. (2007). The employment selection interview: Disparity among, research-based recommendations, current practices and, what matters to Human Rights Tribunals. *Canadian Journal of Public Administration, 24*, 30–44.

Tajfel H, Turner JC. (1986). The social identity theory of inter-group behavior. In Wrochel S, Austin WG (Eds.), *Psychology of Intergroup Relations* (pp. 7–24). Chicago: Nelson-Hall.

Tetlock PE, Boettger R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology, 57*, 388–398.

Tetlock P, Mitchell G, Murray TL. (2008). The challenge of debiasing personnel decisions: Avoiding both under and overcorrection. *Industrial and Organizational Psychology, 1*, 439–443.

Tsui AS, Egan TD, O'Reilly CA. (1992). Being different: Relational demography and organizational attachment. *Administrative Science Quarterly, 37*, 549–579.

Tsui AS, O'Reilly CA. (1989). Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *Academy of Management Journal, 32*, 402–423.

Waldman DA, Avolio BJ. (1991). Race effects in performance evaluations: Controlling for ability, education, and experience. *Journal of Applied Psychology, 76*, 897–901.

Walsh JP, Weinberg RM, Fairfield ML. (1987). The effects of gender on assessment centre evaluations. *Journal of Occupational Psychology, 60*, 305–309.

Wesolowski MA, Mossholder K. (1997). Relational demography in supervisor–subordinate dyads: Impact on subordinate job satisfaction, burnout, and perceived procedural justice. *Journal of Organizational Behavior, 18*, 351–362.

Wessel JL, Ryan AM. (2008). Past the first encounter: The role of stereotypes. *Industrial and Organizational Psychology, 1*, 409–411.

Wiesner WH, Cronshaw SF. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology, 61*, 275–290.

Wiley MG, Eskilson A. (1985). Speech style, gender stereotypes, and corporate success: What if women talk more like men? *Sex Roles, 12*, 993–1007.

Williamson LG, Campion JE, Malos SB, Roehling MV, Campion MA. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology, 82*, 900–912.