



Developing the Research Basis for Controlling Bias in Hiring

Marc Bendick, Jr.*

Bendick and Egan Economic Consultants, Inc.

Ana P. Nunes

Institute of Social Sciences of the University of Lisbon

Discrimination based on race, ethnicity, national origin, gender, age, disability, gender orientation, and other characteristics continues to distort employers' hiring decisions and thereby limit employment opportunities for historically excluded groups. Research in psychology, sociology, economics, and management provides insights concerning the mechanisms of bias and interventions to mitigate their effects, but important questions remain. The innovative research technique of matched pair testing offers laboratory-like controlled conditions in quasi-experiments in real-world hiring situations. We propose seven types of testing studies to advance conceptual understanding of hiring bias and improve hiring practices.

Employers in America's private sector decide who to hire 60 million times a year—more than 250,000 times each work day (U.S. Bureau of Labor Statistics, 2009). For successful job candidates, the hiring process provides employment and determines initial job titles, work assignments, and wages, which, in turn, often influence that employee's career for years thereafter. For successful and unsuccessful job applicants, it also provides job-seeking experience, career information, and encouragement or discouragement. Few human resource management processes rival hiring in impact on the distribution of employment opportunities and rewards.

*Correspondence concerning this article should be addressed to Marc Bendick, Jr., Bendick and Egan Economic Consultants, Inc., 4411 Westover Place NW, Washington, DC 20016 [e-mail: bendickegan@mindspring.com].

The authors contributed equally to this article.

An earlier version was presented at the Monash University Conference on Field Experiments on Discrimination in Markets, Prato, Italy (2005).

Controlling bias that potentially distorts these high-stakes decisions has long been a goal of American society as reflected in federal, state, and local laws against hiring discrimination and requiring affirmative action. It has also motivated many employers to adopt highly structured hiring procedures (Bielby, 2000), train hiring decision makers (Bendick, Egan & Lofhjelm, 2001), and actively “manage workforce diversity” (Kochan et al., 2003).

Drawing primarily on studies in the United States and other industrialized economies, Section I of this article documents that research in psychology, sociology, economics, and management offers many insights into the mechanisms of bias and the likely interventions to mitigate their effects. However, as Sections II and III document, despite considerable progress, a mix of covert and overt bias continues to pervade the American hiring system. Accordingly, this article proposes additional research to support further changes in hiring practices. Section IV proposes four types of studies to expand conceptual understanding of hiring bias, while Section V proposes three types to enhance the influence of such research on hiring practitioners.

Other publications synthesize research on employment bias and identify future research needs (e.g., Brief, 2008; Dipboye & Colella, 2005; Stockdale & Crosby, 2004). The present article differs from these by focusing on hiring and one innovative research methodology—matched pair testing—with unique potential for studying hiring. This technique is described in Section III.

Section I: Research-Based Predictions about Hiring Bias

The Civil Rights Movement of the 1960s was a reaction to widespread, blatant, and sometimes legally codified discrimination against African Americans and other groups. Correspondingly, ensuing antidiscrimination legislation—prominently, the Civil Rights Act of 1964 (mandating equal opportunity), Executive Order 11246 (establishing affirmative action), the Age Discrimination in Employment Act, the Americans with Disabilities Act, and counterpart state and local laws—aimed foremost at employers’ *conscious* behavior. Violations of these statutes are most commonly proven through documentation of discriminatory acts (e.g., use of racial epithets in the workplace) or measurement of employment outcomes (e.g., few minority employees when many qualified minority job applicants are available).

Behavioral science research raises important concerns about the limitations of this approach. In particular, it questions the effectiveness of seeking to change employer behavior without explicitly addressing the often-unconscious attitudes and perceptions underlying that behavior (Gaertner & Dovidio, 2005). Equally, it suggests the need to improve specific employment processes (such as interviewing, performance evaluation, or succession planning), not simply the employment

outcomes they help to shape. At least three lines of research raise such concerns explicitly or implicitly.

Stereotypes Infect Us All

The first line of research explores stereotypes and their role in discriminatory behavior. *Implicit bias*—unconscious association of traits with members of a demographic group (Greenwald, McGhee, & Schwartz, 1998; Nosek, Greenwald, & Banaji, 2005)—has been demonstrated and these associations shown to correlate with biased behavior (Dovidio, Kawakami, & Gaertner, 2002). When these associations are activated in the hiring process, their predominantly negative content about traditionally excluded groups (e.g., African Americans are uneducated; women are not career-committed) handicap members of these groups in competing for jobs.

Decades of related research have further elucidated how stereotypes unconsciously influence perceptions and evaluations, and process central to employment decisions such as hiring. For example, studies have documented that in evaluating members of a stereotyped group, individuals pay more attention to information consistent with a stereotype than to inconsistent information (Koomen & Dijkster, 1998), interpret ambiguous information to confirm stereotypes (Hilton & Van Hippiel, 1996), seek information confirming stereotypes at a greater rate than that contradicting them (Erber & Fiske, 1984), and are unaffected by information that a stereotype is invalid (Nelson, Acker, & Manis, 1996). Individuals also make memory errors consistent with stereotypes (Eberhardt, Dasgupta, & Banaszynski, 2003), even when recalling objective facts such as test scores (Darley & Gross, 1983). Together, such processes explain how hiring decision makers may honestly perceive themselves as making unbiased decisions reflecting objective differences in applicants' qualifications when, in fact, they have not.

Individuals Cannot Readily Counter Stereotypes

A second line of research demonstrates the difficulties that stigmatized groups face when attempting to ameliorate the adverse effects of negative stereotypes. For example, when an individual performs in a way inconsistent with a stereotype, that performance gets discounted as reflecting exceptional circumstances such as luck (Swim & Sanna, 1996). Furthermore, social congruity theory (Eagly & Karau, 2002; see also Fiske, Bersoff, Borgida, Deaux, & Heilman, 1991) predicts that members of a stereotyped group who behave or occupy social roles inconsistent with a stereotype experience stronger adverse reactions than if they conform.

Compounding these difficulties, biased treatment itself or anticipation of it, can adversely affect the actual performance of employees or job applicants. For example, Word, Zanna, & Cooper (1974; see also Shelton, Richeson, &

Salvatore, 2005) demonstrated that White job interviewers sat further away from African American job applicants than White applicants, made less eye contact, and made more speech errors during interviews, and that this interviewer behavior caused applicants to perform less well during interviews. That is, these situations may elicit *stereotype threat* (Roberson & Kulik, 2007; Steele & Aronson, 1995) leading minorities and women to perform less well when they are aware that their performance may confirm a negative stereotype (e.g., when gender is made salient prior to a math exam).

Shared Traits Confer Advantage

A third line of research focuses on *in-group bias*, or the tendency for individuals to prefer members of their own group or derogate those of other groups, as part of maintaining a positive social identity (Brewer, 1979; Tajfel, 1982). Operating consciously or unconsciously, this process then leads to employment decisions not based on applicants' abilities to perform the job, but rather categorization. Its presence has been empirically demonstrated in the laboratory with even minimal or fabricated groups (Hertel & Kerr, 2001; Tajfel, Billig, Bundy, & Flament, 1971). The same processes are likely to operate even more powerfully when based on more salient traits such as race, gender, age, or social class. For instance, Finkelstein, Burke, and Raju (1995) concluded that young persons with authority to hire tended to rate young job applicants more favorably than older ones.

Employment advantages reflecting shared traits may also be created by social patterns outside the workplace. Many relationships in American society today remain highly segregated, with racial, ethnic, and class groups tending to reside in separate neighborhoods and attending different schools and churches. Different genders and ages tend to socialize in different groups and social networks (Alba, Logan, & Stutts, 2000). But social relationships created through such interactions are resources in the labor market (Adler & Kwon, 2002; Granovetter, 1995; Ibarra, 1995). Many job vacancies are never publicly advertised in newspapers or on Internet job boards, so they become known primarily to the friends, relatives, neighbors, classmates, or colleagues of current employees (Kuhn & Skuterud, 2000). Similarly, in preemployment skill testing and interviewing, informal coaching and insider information may equip applicants who have ties to current employees to perform better than their competitors (Hulett, Bendick, Thomas, & Moccio, 2008). Limited access to information-rich social networks helps to keep traditionally excluded groups excluded.

Bias May Evolve Rather than Disappear

Over the four decades since the Civil Rights Movement, antidiscrimination legislation has reduced many blatant forms of employment discrimination, such

as the once-traditional advertising of job vacancies as “Help Wanted-Male” or “Help Wanted-Female.” Over the same period, public opinion polls reflected downward trends in individuals’ self-reported prejudice. For example, in 1965, 59% of Americans agreed that they would vote for a qualified African American presidential candidate, but by 2005, that figure had risen to 93%; parallel questions concerning a qualified woman elicited 55% agreement in 1965 but 87% in 2005 (Gallup, 2005; see also Smith, 2000).

However, these developments may not translate into substantially reduced bias. Rather, blatant discriminatory behavior may have simply evolved into what McConahay, Hardee, and Batts (1981) label *modern racism*—a shift in social norms making explicit expression of prejudiced attitudes rarely seen yet remaining in more covert forms.

Consistent with this model, researchers have documented covert prejudice in the behavior of actual supervisors and managers. For example, Moss & Tilly (2002) and Rooth (2007) describe how, in rejecting minority job applicants, hiring decision makers now often cite applicants’ deficiencies in “soft skills” (“inability to communicate,” “lack of work commitment”) as a socially acceptable proxy for an applicant’s race. Moss and Tilly’s work is consistent with Dovidio and Gaertner’s (2000) findings that discrimination against minorities is most evident when qualifications are ambiguous rather than very strong or very weak; ambiguous qualifications allow bias to influence hiring decisions while leaving hiring decision makers feeling justified in their choices. Other researchers have described some employers’ equal employment initiatives as primarily symbolic, appearing to comply with legal and social norms without intending to change employment outcomes substantially (Bendick & Egan, 2009; Edelman & Petterson, 1999).

Hiring Is Particularly Vulnerable to Bias

Of course, such cognitive and organizational processes can lead to biased outcomes in *post*-hiring employment processes such as performance evaluations, promotions, raises, and terminations. However, hiring decisions differ from post-hiring decisions in ways that tend to make bias more influential in hiring than in those other processes.

The first way is the limited information on which hiring decisions are based (Altonji & Pierret, 2001). Post-hiring decisions tend to be made by supervisors who have observed employees’ performance over months or years. Hiring, in contrast, typically requires assessing job applicants who are virtual strangers. Resumes, applications, and work samples typically provide only a sketch of their qualifications and require judgment about relevance to the positions for which they are being hired. Job interviews tend to be brief—for entry-level positions, often as short as 10 minutes (Bendick, Rodriguez, & Jayaraman, 2010). Interviews also tend to be subject to “impression management,” deliberate self-presentation

to create impressions not sustained post-hiring (Giacalone & Rosenfeld, 1989; Higgins & Judge, 2004). For these and other reasons, traditional job interviews have limited power to predict post-hiring job performance (Le, Oh, Shaffer, & Schmidt, 2007).

Another factor negatively impacting stigmatized groups is the time pressure under which managers often operate. Stereotypes exercise particular influence in time-pressured situations (Macrae, Bodenhausen, & Milne, 1998). Concurrently, limited contact between job applicants and hiring decision makers imbues every interaction with intense pressure to perform that tends to enhance stereotype threat.

Another circumstance enhancing bias in hiring is differences in opportunities to correct errors. Many post-hiring decisions are part of ongoing or repeated processes in which errors can be corrected later. For instance, a too-generous raise in one year may be brought back into line by a smaller raise the next year, or employees passed over for one promotion may be considered for others. In fact, the continued visibility of minority employees who are not progressing may itself pressure managers to act (Stangor, Sechrist, & Jost, 2001; Zitek & Hebl, 2007). In hiring, in contrast, applicants who are not hired tend immediately to go elsewhere to find employment, eliminating opportunities for employers to reconsider.

External pressure from antidiscrimination laws is also less likely to correct bias in hiring decisions than post-hiring decisions. Employees who feel aggrieved about a post-hiring decision often have information on which to base a complaint (e.g., who was promoted instead of me?) and a sufficient stake in the outcome to seek redress. In contrast, job applicants who suspect discrimination often lack sufficient information to determine if discrimination has occurred (e.g., when the company said the job vacancy was already filled, was that an excuse?), and are more likely to pursue other opportunities than to dispute the decision. For such reasons, hiring complaints comprise only 6% of formal discrimination complaints filed with the Equal Employment Opportunity Commission (EEOC) (Bendick, Jackson, & Reinoso, 1994).

Section II: Empirical Evidence of Employment Bias

The research reviewed in Section I translates into a testable hypothesis: bias continues to adversely affect hiring outcomes for historically excluded groups in the American labor market today. Does empirical evidence support this hypothesis?

Certainly, research documents substantial progress toward employment equality over recent decades. Women now constitute 51% of managerial and professional workers in the American labor force, and race/ethnic minorities more than 22% (U.S. Bureau of Labor Statistics, 2008), a very different situation from the early 1960s, when their numbers in most such occupations were so small that essentially every individual was a highly visible exception. Women's earnings,

which averaged about 60% of men's until the 1970s, rose to nearly 80% by the 1990s (Blau & Kahn, 2007), and over the same period, African Americans' average earnings rose from 57% of Whites' to more than 73% (Smith & Welch, 1989).

Yet, such measures document remaining problems as much as progress. Earnings ratios of 73% or 80% are still far below the 100% that would signal simple equality. Worse, in recent years, upward movement of race and gender wage ratios has slowed or stopped (Rodgers, 2006). Concurrently, the number of women and minorities remains very limited in many occupations, especially more prestigious, well paid ones—patterns of occupational segregation often referred to as “glass walls” and “glass ceilings” (Altonji & Blank, 1999; Reskin & Bielby, 2005). For example, although Hispanics now constitute 11.6% of police patrol officers, they are only 6.1% of police sergeants and lieutenants. African Americans are 7.7% of construction laborers but only 2.2% of structural steel workers (U.S. Bureau of Labor Statistics, 2008; see also Bendick, 2000). And although women now are 31% of medical doctors, they are only 4% of orthopedic surgeons. Moreover, female physicians earn an average of 18% less than male physicians with matching credentials, medical specialties, years in practice, and work hours per week (HRSA, 2010).

Research in the form illustrated by these findings on physicians' earnings—showing different employment outcomes among demographic groups not explained by differences in education, experience, other measures of qualifications, or work effort—provides the most rigorous evidence that substantial bias remains in the American labor market. Such studies have been conducted on many different occupations, industries, localities, and demographic groups, producing results such as the following:

- After controlling for education, age, work experience, residence, and criminal records, dark-skinned black males had a 52% lower chance of being employed than lighter skinned black males (Johnson, Bienenstock, & Stoloff, 1995).
- Although surgery did not alter their job qualifications, when transgender individuals who were men transformed into women, their earnings fell 12%, while women who transformed into men had their earnings rise 8% (Schilt & Wiswall, 2008).
- Among American professionals and managers working in international business, an additional year of international experience raised men's annual earnings \$2,500 but women's only \$1,300; working an extra 8 hours a week increased men's earning \$7,300 but women's only \$4,300 (Egan, Bendick, & Miller, 2002).
- Leading symphony orchestras auditioning musicians behind screens so that judges could not see the applicants hired more women players than those conducting auditions in the open (Golden & Rouse, 2000).

Such findings are consistent with perceptions of bias reported by adversely affected individuals. In one typical survey with a nationally representative sample, 81% of African Americans, 60% of Hispanics, and 53% of Asian respondents felt that they would have a lower chance of promotion to a managerial position than an equally qualified White (Smith, 2000). In another national survey, 31% of Asians, 26% of African Americans, 18% of Hispanics, and 22% of women reported having been discriminated against in their workplace during the previous year (Gallup, 2005). In narrower examples, in a survey of self-identified transgender adults in San Francisco, 40% of respondents reported having been discriminated against in applying for jobs, and 24% reported having been sexually harassed at work (Guardian, 2006); and in a nationwide survey of women firefighters, 85% reported having been treated adversely at work, including hostile comments, sexual advances, and being placed in unnecessary danger (Hulett et al., 2008).

Finally, evidence of perceived discrimination is provided by formal legal complaints filed with the United States, which in 2008 numbered 93,000 (U.S. EEOC, 2010). In parallel, thousands of antidiscrimination lawsuits by public antidiscrimination agencies or private litigators continue to be filed each year, with some resolutions including tens of millions of dollars in damages (Bendick & Egan, 2009; Darity & Mason, 1998).

Section III: Testing Provides Direct Evidence of Hiring Bias

Although the previous section provides considerable evidence of continuing bias, this evidence often does not separate hiring from post-hiring processes; may be based on perceptions rather than objectively verified; requires extrapolation from laboratory settings to behavior of actual employment decision makers; or is inferred from differences in employment outcomes that remain after other explanations have been eliminated. Since 1990, an additional research technique has been increasingly used that uniquely provides *direct, objective* observation of *hiring bias in real-world* settings under controlled, quasi-experimental conditions. This technique is matched pair testing.

Matched pair testing (also referred to as situation testing, paired comparison testing, employment auditing, field experiments, or employment testing) is a systematic procedure creating controlled experiments analyzing employers' candid responses to employees' personal characteristics (Bendick, 2007; Pager & Western, 2012). Economists define employment discrimination as valuation in the labor market of workers' characteristics not related to productivity (Arrow, 1998). In testing, pairs of research assistants apply for the same actual job vacancy. Within each pair, applicant characteristics related to a worker's productivity on the job—education, work experience, professional certifications, and technical skills—are equalized by selecting, training, and credentialing testers

to appear equally qualified for the positions they seek. Simultaneously, personal characteristics unrelated to job performance are experimentally manipulated by pairing testers who differ in one characteristic—a White paired with an African American, a male with a female, or a person age 32 with one age 57. If testers within a pair experience substantially different responses to their job-seeking efforts, it is then presumed that this differential treatment is due to the employers' reaction to the manipulated characteristic. Thus, testing addresses Kang's (2012) *Axis of Specificity* by documenting both specific discrimination perpetrated by an employer that may be used in litigation and general information regarding levels of discrimination in broader circumstances that may be used to develop public policy.

Of course, interpretation of testing outcomes as discrimination is appropriate only if employers are presented with pairs of job candidates who truly appear equally qualified. This condition is relatively easy to achieve in testing that involves resumes mailed, faxed, or e-mailed to employers (Bendick, Jackson, & Romero, 1996; Bertrand & Mullainathan, 2004). In these studies, the resumes describe equivalent education, work experience, and job skills while varying formats and details to avoid appearing obviously similar. The resumes communicate applicants' demographic characteristics through gender-specific names, ethnically related activities, or age-revealing graduation dates. In nations where resumes customarily include applicant photographs, these images can also communicate demographic characteristics.

Resume testing can probe only the initial stages of the hiring process, ending with employers' decision to invite candidates for in-person interviews. To study the complete hiring process, "live" testers are required to file applications, return messages, take skills tests, and be interviewed. Ensuring that the testers in each pair appear equally qualified throughout those processes requires substantial care and effort (Bovenkerk, 1992; Lodder, 1994).

The first step is to recruit research assistants who meet a daunting set of requirements: ability to play the job seeker role convincingly while accurately observing the hiring process; willingness to approach the study objectively; similarities between testing partners in general appearance and demeanor; and the demographic characteristics required by the study design. Recruiting individuals meeting all these requirements is often a time-consuming, painstaking process; in one typical study, 93 potential testers were interviewed before four were selected to form two testing teams (Nunes & Seligman, 1999). College students, professional actors, actual job seekers, and community volunteers have all served as testers.

The second step is training, which typically requires at least three days, to make pairs of testers equally credible job applicants. During training, testers develop and memorize their false resumes, receive coaching on effective interviewing techniques, and rehearse similar answers to common interview questions.

Concurrently, testers are trained to remember important details of their testing experiences.

A third step in maintaining tester equivalence involves closely supervising testers' actions. The testers within each pair usually present themselves to employers in random order, with the second tester applying shortly after the first. Testers document their experiences as soon as practical after each event without knowing the experiences of their testing partners. They typically use structured questionnaires and are constantly reminded to focus on facts rather than interpretations. Such careful management requires continuous, hands-on monitoring by a trained "Test Coordinator," who can usually supervise no more than three teams concurrently.

Testing studies typically repeat their quasi-experiment for dozens or hundreds of job vacancies, to "average out" random circumstances that may affect the outcome in single tests. In analyzing outcomes, one key summary statistic is the *net rate of discrimination*, the proportion of job applications in which testers with the characteristic hypothesized to be disfavored is successful minus the proportion of applications in which testers with the characteristic hypothesized to be favored is successful. Another important measure is the ratio of the proportion of tests in which testers with the disfavored characteristic are successfully divided by the proportion of tests in which the other tester is successful. "Successful" is typically defined as reaching an identifiable milestone in the hiring process, such as being offered an interview or a job.

Given the 20–40% net rates of discrimination observed in typical testing studies, statistically significant main effects in the experiment can be obtained with as few as 40–50 completed tests. Samples of about 100 tests have proved sufficient to observe statistically significant effects of mediating factors on net rates of discrimination. Analyses applying multivariate techniques, such as regression analysis (Kenney & Wissoker, 1994), often require larger samples.

The following are examples of hiring outcomes that testing studies have identified as biased:

- A major newspaper carried an advertisement for a restaurant supervisor in an affluent neighborhood. An African American tester who presented himself at the restaurant was told that he would be called if the restaurant wished to pursue his application. Minutes later, a White tester whose resume showed the same level of education and restaurant experience followed the same procedure. He was called later that day to schedule an interview, and subsequently offered the position. The African American tester made four follow-up calls to reiterate his interest, including shortly after the White tester declined the job offer, with no response (Bendick et al., 1994).

- A vacancy for a receptionist in an optometrist's office was advertised in a local newspaper in an affluent neighborhood. When a tester with a Latina name and slight accent telephoned the following day, she was put on hold, called Carmen when she had given her name as Juanita, and told that the office was not taking any further applications. When her testing partner with an Anglo name and no accent called 13 minutes later, she received an interview appointment for the following morning (Bendick, Jackson, Reinoso, & Hodges, 1991).
- An employment agency advertised for an executive recruiter. Two White males, whose resumes and appearance portrayed ages 32 and 57, respectively, responded by telephone and were granted interviews. The older tester's interview lasted 48 minutes, during which he was cautioned against making a precipitous career change and instructed to call back if he was still interested after reading books on sales techniques. The younger tester's interview lasted 85 minutes, with the interviewer discussing work and nonwork topics in a friendly manner and commenting enthusiastically on the tester's responses. This tester was invited back for a second interview, after which he was offered a job (Bendick, Brown, & Wall, 1999).
- An automobile service shop advertised in a newspaper for a technician to lubricate and repair automobiles. When a female applicant whose resume showed experience in physically demanding jobs applied for the position, the manager told her that "the auto lube job is hard for a woman," said that he liked her smile, and offered a lower paying position in the on-premise coffee kiosk. When her male testing partner applied several hours later, he was interviewed for the advertised technician position (Nunes & Seligman, 2000).

The most frequent criticism of testing studies is by "free market" economists who reason that, because employers are forced by a competitive labor market to consider only productivity-related characteristics of job applicants, differences in hiring outcomes between paired testers must signal failure by the researchers to match the testers on some subtle productivity-related characteristics (e.g., Heckman, 1998; see also Pager, 2007). Such criticism is tautological as well as contradicted by research reviewed throughout this article. However, it provides an important caution that only researchers committed to methodological rigor should undertake testing.

Over the past two decades, several dozen testing studies have examined hiring in labor markets from Boston to Los Angeles and occupations from entry-level retail sales to professional and managerial positions (Bendick, 1999; Bendick, 2007; Pager, 2007). Several dozen additional studies have been conducted in other industrialized nations, from the Netherlands to Australia (European Commission, 2006; ILO, 1998; Riach & Rich, 2002). Demographic groups whose experiences

were analyzed have included women, older workers, persons with disabilities, transgender individuals, and race/ethnic minority groups ranging from African Americans and Hispanics in the United States to immigrant Turks, North Africans, West Indians, and South Asians in Europe.

These studies have been essentially unanimous in documenting considerable hiring bias. As mentioned previously, typical net rates of discrimination range between 20% and 40%. Equivalently, these studies estimate that bias infects the hiring decisions of 20–40% of employers. Such estimates are striking consistent with the nontesting evidence reviewed in Sections I and II.

One obvious direction for future testing is to continue “mapping” the prevalence of discrimination in different locations (urban centers vs. suburban areas), occupations and industry (professional occupations vs. entry-level jobs), and demographic groups (women vs. African Americans). Such studies would be particularly useful if they were repeated using a consistent methodology on a nationally representative sample to generate a recurrent “national report card” tracing trends over time (Fix & Turner, 1999). Another particularly useful form of “mapping” involves testing for bias on bases other than the “usual” race, gender, and age characteristics; innovative studies have involved, for instance, Arab Americans facing backlash after September 11, 2001 (Discrimination Research Center, 2004), persons with a criminal record (Pager & Western, 2012), transgender individuals (Make the Road, 2010), and overweight persons (Rooth, 2009).

However, testing has typically revealed relatively modest variation in the prevalence of discrimination among demographic groups, locations, and occupations, suggesting that exploration of how rates of discrimination vary across demographic groups or labor markets should not command high priority. Furthermore, although these studies typically capture media attention when released, there is little evidence that they have powerful or lasting impact on public opinion or public policy. For example, although testing results were discussed in debates about California’s Proposition 209 to abolish affirmative action (Bendick, 1995), the proposition passed.

Accordingly, matched pair testing is likely to address bias more effectively if, rather than primarily measuring the prevalence of hiring bias, studies seek to advance fundamental understanding of bias and its remedies. Compared to past testing studies, research in this spirit would be more grounded in behavioral science theory, coordinated with nontesting research on the same issues, and designed to provide data analyzable with other methodologies. In short, future research should “test deeper” rather than “test broader.”

Section IV: Testing to Understand Hiring Bias

This section presents four research proposals illustrating that “test deeper” approach.

Study Employer–Job Candidate Interactions

The first proposal concerns using testing to analyze employer–candidate interactions, particularly job interviews. Behavioral science research suggests that bias, especially when unconscious, is often embodied in small, subtle but crucial differences in words or actions referred to as *microinequities* (Valian, 1998). These small differences often leave perpetrators of bias unaware of these unconscious processes and their cumulative impact. Equally, they pose measurement challenges to researchers seeking to understand exactly how bias operates.

To date, research on the details of employer–employee interactions has primarily examined written materials such as letters of recommendations and performance evaluations (Bison-Rapp, 1999; Trix & Psenka, 2003). These studies have documented systematic differences in statements about equally qualified individuals of different demographic backgrounds. For example, a comparison of performance evaluations for men and women professionals who all received high-performance ratings at a financial services firm found that women were praised primarily for activities within their own work group and men for external activities; criticisms of men tended to be accompanied by mitigating explanations but those for women were not; and men were commonly recommended for advancement, while women were described as valuable in their current position (Townsend, 1997). Other studies have revealed *Linguistic Intergroup Bias*, language describing in-group and out-group behavior that itself perpetuate stereotypes—for instance, by describing out-group positive behavior and in-group negative behavior in concrete terms suggesting situational traits, while reporting out-group negative behavior and positive in-group behavior in abstract terms suggesting persistent qualities (Maas, Salvi, Acuri, & Semin, 1989; Semin & Fiedler, 1992).

Only a few studies in this tradition have examined face-to-face interactions rather than written materials (Binning, Goldstein, Garcia, & Scatterregia, 1988; Shelly & Shelly, 2009). This research requires samples of the exact words used to describe demographically different individuals who are equivalent in actual qualifications or performance—precisely what testing can provide. Through miniature technology, voice and video recordings of conversations can be collected without making employers aware of being tested or recorded. Transcripts of these encounters could then be analyzed using psycholinguistic techniques.

Some testing studies have examined employer and employee behavior in job interviews. For example, Bendick et al. (1994) analyzed interviews of African American and White applicants for entry-level positions. They observed that although the two groups received very different numbers of job offers, applicants experienced no substantial differences in treatment during interviews such as interview length, proportion of the interview devoted to job-relevant topics, and rank of the interviewer. Another study examined interviews of Whites and Persons of Color applying for waitstaff positions in upscale restaurants (Bendick, Rodriguez,

& Jayaraman, 2010). It documented that interviewers tended to accept White applicants' claims of past restaurant experience without probing but skeptically questioned non-Whites. Such studies suggest the insights that more sophisticated analyses of interview transcripts could uncover.

Coordinate Testing Studies of Multiple Out-Groups

Our second research proposal is to study patterns of bias against multiple out-groups by the same employer.

Many analyses of discrimination focus on issues specific to demographic groups, for example, the residual effects of segregation on African Americans, the relationship between family responsibilities and women's careers, or the effect of English language requirements on recent immigrants (Bell, 2007). But other threads in behavioral science research suggest a fundamentally different approach. More than 50 years ago, psychologists began to consider prejudice as a consistent characteristic of an individual so that a person harboring bias against women would also tend to be prejudiced against minorities (Bierly, 1985; Peterson, Doty, & Winter, 1993). In parallel, some sociologists have argued that the essence of workplace discrimination is not adverse attitudes toward specific out-groups but rather *social closure* to preserve the power, status, and privilege of a dominant in-group against *all* alternative claimants (Freshman, 1990; Moore, 1990; Murray, 1988; Sidanius & Pratto, 2001; Tomaskovic-Devey, 1993).

This debate is important not only to social theory but to practical antidiscrimination efforts. The first perspective implies that it is important to contradict stereotypes adverse to individual out-groups. In contrast, the second perspective considers stereotypes primarily *ex post* rationales for negative treatment of out-groups rather than independent causes and suggests controlling the generic process of stereotyping rather than the content of specific stereotypes (Bendick et al., 2001; Egan & Bendick, 2008).

To date, testing has typically examined one out-group per study. However, studies could be organized to, for example, field teams of males and females, African Americans and Whites, and older and younger workers to apply to the same employers. Positive correlations in bias against multiple out-groups would support the social closure model, while zero or negative correlations would support the alternative (Fiske, Cuddy, Glick, & Xu, 2002).

Behavioral science research also suggests other ways in which an employer's behavior toward an out-group may differ depending on a context that includes other out-groups. For instance, *moral credentialing* (Monin & Miller, 2001) suggests that employers who have hired an applicant from one stigmatized group, and therefore feel that they have adequately demonstrated egalitarian values, may be less favorable to subsequent candidates from other stigmatized groups. Again, testing studies could be designed to test this hypothesis, in this case, by appropriately

sequencing multiple tests of the same employer and analyzing the relationship between each test outcome and that of preceding tests.

Evaluate Strategies for Minimizing Bias

Practical advice routinely offered to out-group job seekers often includes suggestions concerning how to minimize bias they may encounter. For example, some advisors counsel explicitly refuting stereotypes about groups to which the job seeker belongs (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000); others suggest using individuating information to counter the applicability of the stereotype to the job seeker (Beckett & Park, 1995; Glick, Zion, & Nelson, 1988); and still others advise emphasizing alternative, positive stereotypes associated with those groups (Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008). Some advisors suggest implementing such strategies proactively—in anticipation that stereotypes inevitably bias hiring decisions—while others suggest doing so only if some indication suggests a problem. Matched pair testing could be designed to provide evidence-based advice to job seekers rather than the current, largely intuitive guidance.

To date, only rare testing studies have examined alternative strategies for combating bias. One study of age discrimination using mailed resumes randomly assigned different cover letters to accompany the older workers' resume. In one letter, older applicants described themselves as career-committed, energetic, and technologically up-to-date—positive attributes stereotypically associated with younger applicants; in an alternative letter, older applicants described themselves as experienced, mature, and stable—positive attributes stereotypically associated with older workers; and a third letter contained neither statement. The first cover letter generated a substantially higher rate of favorable employer responses than the other two (Bendick et al., 1996).

Employers also have implemented a variety of initiatives designed to minimize bias. Most prominently, the majority of U.S. employers today invest in “workforce diversity training” for their employees, despite research that questions the effectiveness of many of these efforts (Bendick et al., 2001; Kalev, Dobbin, & Kelly, 2006). Testing could be structured to provide assessments of these efforts. For example, tests could be conducted in a workplace prior to a diversity training program and then again subsequently. Or a multiestablishment firm could implement training in one subset of its establishments and not in a matched subset, and then conduct hiring tests in both experimental and control locations. Here, testing could measure the effect of the training on overall rates of discrimination and specific hiring practices that employees were trained to adopt or avoid.

Testing could also improve guidance for employers on other antibias efforts. For example, in sincere efforts to avoid violating social norms, some hiring interviewers consciously avoid any reference to race. However, Apfelbaum, Sommers,

and Norton (2008) demonstrate that such *strategic colorblindness* may backfire, leading to nonverbal unfriendliness interpreted by Black observers as negative and prejudiced. Similarly, Chartrand and Bargh (1999) discuss a *chameleon effect* in which interviewees unconsciously mirror the behavior of interviewers, generating awkward behavior not representative of their potential post-hiring performance. Testing could be used to study the prevalence in actual job interviews of these hypothesized effects and whether they vary among different demographic combinations of interviewers and interviewees, and findings could be translated into practical guidance to interviewers.

Study the Effects of Hiring Rejection

Research has long documented the negative psychological and social consequences of unemployment, including increased physical and mental illness, loss of self-esteem, family stress, and discouragement (Clark, Georgellis, & Sanfey, 1999; Linn, Sandifer, & Stein, 1985). A few studies have focused on these effects when bias was a factor (Goldsmith, Sedo, Darity, & Hamilton, 2004). These studies employ a range of techniques, including epidemiological studies of morbidity and mortality, surveys utilizing scales of self-efficacy and psychological well-being, laboratory studies tracking physiological responses to disappointment, and diaries tracking attitudes and perceptions over time (Bolger, Davis, & Rafielli, 2003).

Matched pair testing offers an additional methodology for measuring these effects. It provides unusually detailed information about the rejection experience. It allows examining hiring separately from other aspects of workers' unemployment experiences, such as the duration of their unemployment. And by forming appropriate tester teams, it can isolate differences in the responses of different types of workers—for instance, youth just entering the workforce compared to mid-career workers.

Using such measures, researchers might study, for example, whether the effects of failing to be hired are different when the decision was biased or unbiased; the cumulative effects of multiple job disappointments; and whether the same disappointing experience creates different effects on different demographic groups. For example, it might be hypothesized that because greater expectations of eventual employment create a greater sense of *self-efficacy* (Bandura, 1997), White males might experience smaller adverse effects than women or minorities. Testing could also be used to examine the efficacy of strategies to minimize the adverse effects of job disappointment. For example, *self-affirmation theory* (Steele, 1988) suggests that individuals frustrated in one endeavor (e.g., job seeking) might preserve their self-image by success in a different domain (e.g., volunteer work). Explicitly measuring testers' perceptions of themselves as well as attitudes toward in-group and out-group members' pre- and posttesting experiences may provide information into how hiring experiences may be internalized differently by

different groups. However, this research would require that testers be actual job seekers who really want jobs and who would be allowed to accept positions offered.

Testing could also be used to study the effect on individuals' attitudes and empathy of being exposed to the reality of discrimination that they otherwise might not encounter. Research (Batson et al., 2003; Gaertner & Dovidio, 2005) suggests that putting an individual "in the shoes of others" expands the range of persons about whose well-being they feel concerned, and Fogelman (1994) has documented the role of striking personal incidents in persuading bystanders to act on behalf of victims of persecution. It can be hypothesized that, in this spirit, White or male testers who observe their testing partners encounter bias may become more empathetic. This hypothesis could be studied by measuring the attitudes and behavior of in-group testers before and after their testing experience.

Section V: Using Testing to Change Employer Behavior

A perennial challenge in the behavioral sciences is to mobilize research findings to influence real-world practices. How can employers be convinced and assisted to incorporate the research findings discussed throughout this article into their firm's employment practices? Testing offers new approaches for doing so, of which this section outlines three proposals.

Testing on Behalf of Employers

Employers commonly monitor their own staff using unobtrusive data gathering techniques. For example, retailers often employ "mystery shoppers" to pose as customers and record their interactions with sales staff (www.mysteryshop.org). The employers use these data to improve customer service. Most employment testing studies have been conducted without employers being aware of being tested or presented with findings about their individual firms. Our first proposal suggests testing by employers as a means of self-audit.

One example of which we are personally aware took place in 2009 at the teaching hospital of a major medical school. This employer desired to offer jobs to residents of its surrounding community as well as employ staff who could relate to its demographically diverse patients, and therefore wanted to ensure that its hiring practices were not biased. To support this goal, the hospital required all supervisors to be trained to conduct job interviews in a job-related structured manner (Le et al., 2007). However, the hospital had never measured the extent to which these supervisors subsequently followed those procedures. Accordingly, the hospital's vice president of human resources secretly arranged with a nonprofit organization to send matched pairs of White and minority applicants to apply for positions at the hospital and two comparable hospitals in the same city.

Businesses currently invest huge numbers of staff hours and many millions of dollars each year to address bias and increase staff diversity. However, 62% of them collect no information on the impact of these efforts (Esen, 2005). This lack of data limits their ability to improve these initiatives as well as to hold managers accountable for following prescribed practices (Bendick, 2008). As the hospital example illustrates, testing conducted by or with employers can provide direct data about the extent to which initiatives effectively modify employment practices “on the shop floor” and the extent to which those modifications change hiring outcomes.

Testing for Training

When testing results are presented to audiences of nonresearchers—for example, in legislative hearings or the mass media—their attention-grabbing power is immediately evident. This power reflects testing’s combination of rigorously controlled statistical evidence and vivid anecdotes that put a human face on those statistics (Cialdini, 2000).

This persuasive power has yet to be extensively harnessed in practical employee training. Among larger employers, more than 90% provide some form of antidiscrimination or prodiversity training, often enrolling everyone from senior executives and mid-level supervisors to nonmanagerial employees (Esen, 2005; Kalev et al., 2006). The most effective forms of this training use vivid examples from real workplaces to persuade trainees of the continued presence of discrimination and communicate desired changes in employee behavior (Adamson, 2000; Bendick et al., 2001). Testing is ideally structured to provide specific vivid examples for use in training, especially if tests are documented through audio or video recording. In addition, posttraining testing could be used to assess whether training is effective.

Testing for Litigation

U.S. employers who violate equal federal or state employment laws can be sued either by government agencies, such as the federal EEOC, or in private litigation brought by the victims of discrimination. If the plaintiffs prevail, the employer may be liable for substantial damage payments as well as mandatory court-supervised changes in employment practices. Since the late 1960s, such litigation—or employers’ desire to avoid it—has been a major motivator for reductions in employment bias (Blumrosen, 1993).

American law grants matched pair testing potential roles in enforcement of these laws. Individual testers and nonprofit organizations employing testers have “standing” to become plaintiffs in litigation based on testing evidence alone (Boggs, Sellers, & Bendick, 1993; Yelnosky, 2010). In addition, if an employer

is sued based on evidence other than testing, testing-based documentation of an employer's discriminatory behavior can be used as corroborative evidence.

Despite this potential, only a handful of testing-based enforcement actions have been brought. One pioneering lawsuit, *Fair Employment Council et al. v. BMC Marketing*, was filed in the District of Columbia in 1990. In it, the plaintiffs were two African American university students and the nonprofit organization that employed them as testers. The defendant was a local office of one of the nation's largest job placement agencies. The agency had interviewed, coached, and found entry-level office jobs for two White testers, while failing to do so for the African Americans who were their testing partners. This litigation was settled with payment of damages to the nonprofit organization and commitments by the defendant to retrain its staff and eliminate discriminatory practices (Boggs et al., 1993).

That lawsuit involved plaintiffs from a stigmatized group—African Americans—for whom testing is only one of several ways to generate evidence of discrimination. For other groups, however, testing-based evidence may be the only feasible basis for litigation. For example, in antidiscrimination litigation, statistics from sources such as the Census are often used to demonstrate that an employer has fewer female or race/ethnic minority employees than would be expected based on their availability in the local labor market. However, such Census data are not collected for many historically excluded groups—for instance, persons with psychiatric disabilities (Tal, Moran, Rooth, & Bendick, 2009) or gay, lesbian, transgender, and bisexual individuals (Make the Road, 2010). In that circumstance, direct evidence of employer discrimination obtained by testing may be the only feasible source of legally viable evidence of hiring discrimination.

Litigation is only one adversarial approach in which testing can be used to influence employers. To date, publicly released results from testing studies have typically not named employers against whom incriminating findings were obtained. However, the results of tests on specific employers could be released to the general public, news media, insurers, investors, unions, customers, and others, providing information by which these stakeholders can induce discriminating employers to change their behavior (Egan, Mauleon, Wolff, & Bendick, 2009).

Section VI: Summary

The seven types of matched pair testing studies discussed here form an ambitious research agenda for any society attempting to reduce bias in hiring. However, these examples by no means exhaust the potential of testing to generate theoretical and practical advances. It is time for researchers in the social, behavioral, and managerial sciences to mobilize the power of this technique creatively to advance this important societal goal.

References

- Adamson, J. (2000). *The Denny's story: How a company in crisis resurrected its good name and reputation*. New York: John Wiley.
- Adler, P., & Kwon, S. (2002). Social capital: Prospects for a new concept. *Academy of Management Review*, 27(1), 17–40.
- Alba, R., Logan, J., & Stutts, B. (2000). How segregated are middle-class African Americans? *Social Problems*, 47(4), 543–558.
- Altonji, J., & Blank, R. (1999). Race and gender in the labor market. In O. Aschenfelder & D. Card (Eds.) *Handbook of labor economics*, (Vol. 1, pp. 3243–3259). Amsterdam: Elsevier.
- Altonji, J., & Pierret, C. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, 116, 313–350.
- Apfelbaum, E., Sommers, S., & Norton, M. (2008). Seeing race and seeming racist? Evaluating colorblindness in social interactions. *Journal of Personality and Social Psychology*, 95, 918–932. doi: 10.1037/a0011990.
- Arrow, K. (1998). What has economics to say about racial discrimination? *Journal of Economic Perspectives*, 12, 91–100.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Batson, C., Lishner, D., Carpenter, A., Dulin, L., Harjusola-Webb, S., Stocks, E., Gale, S., Hassan, O., & Sampat, B. (2003). “As you would have them do unto you”: Does imagining yourself in the other’s place stimulate moral action? *Personality and Social Psychology Bulletin*, 29, 1190–1201. doi: 10.1177/0146167203254600.
- Beckett, N., & Park, B. (1995). Use of category versus individuating information: Making base rates salient. *Personality and Social Psychology Bulletin*, 21, 21–31. doi: 10.1177/0146167295211004.
- Bell, M. (2007). *Diversity in organizations*. Mason, OH: Thomson South-Western.
- Bendick, M., Jr. (1995). Research evidence on racial/ethnic discrimination and affirmative action in employment. In *Discrimination and affirmative action: Are there any facts out there?* (pp. A58-A78). Sacramento: Assembly Judiciary Committee, California State Legislature.
- Bendick, M., Jr. (1999). Adding testing to the nation’s portfolio of information on employment discrimination. In M. Fix and M. Turner (Eds.) *A national report card on discrimination: The role of testing* (pp. 47–68). Washington, DC: The Urban Institute.
- Bendick, M., Jr. (2000). Using EEO-1 data to analyze allegations of employment discrimination. *Presented at American Bar Association National Conference*.
- Bendick, M., Jr. (2008). Measure inclusion, not diversity. *Presented at Society for Human Resource Management National Diversity Conference*.
- Bendick, M., Jr., Brown, L., & Wall, K. (1999). No foot in the door: An experimental study of employment discrimination against older workers. *Journal of Aging and Social Policy*, 10, 5–23.
- Bendick, M., Jr., & Egan, M. (2009). *Research perspectives on race and employment in advertising*. Washington, DC: Bendick and Egan Economic Consultants, Inc. for the Madison Avenue Project.
- Bendick, M., Jr., Egan, M., & Lofhjelm, S. (2001). Diversity training: From anti-discrimination compliance to organization development. *Human Resource Planning*, 24, 10–25.
- Bendick, M., Jr., Jackson, C., Reinoso, V., & Hodges, L. (1991). Discrimination against Latino job applicants: A controlled experiment. *Human Resource Management*, 30, 469–484.
- Bendick, M., Jr., Jackson, C., & Reinoso, V. (1994). Measuring employment discrimination through controlled experiments. *Review of Black Political Economy*, 23, 25–48.
- Bendick, M., Jr., Jackson, C., & Romero, J. (1996). Employment discrimination against older workers: An experimental study of hiring practices. *Journal of Aging and Social Policy*, 8, 25–46.
- Bendick, M., Jr., Rodriguez, R., & Jayaraman, S. (2010). Race-ethnic employment discrimination in upscale restaurants: Evidence from paired comparison testing. *Social Science Journal*, 47, 802–818.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Brendan more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 113, 991–1001.

- Bielby, W. (2000). Minimizing workplace gender and racial bias. *Contemporary Sociology*, 29, 120–129.
- Bierly, M. (1985). Prejudice toward contemporary outgroups as a generalized attitude. *Journal of Applied Social Psychology*, 15, 189–199. doi: 10.1111/j.1559-1816.1985.tb02344.x.
- Binning, J., Goldstein, M., Garcia, M., & Scatterregia, J. (1988). Effects of preinterview impressions on interviewer questioning strategies in same- and opposite-sex employment interviews. *Journal of Applied Psychology*, 73, 30–37. doi: 10.1037/0021-9010.73.1.30.
- Bison-Rapp, S. (1999). Bulletproofing the workplace: Symbol and substance in employment discrimination law practice. *Florida State University Law Review*, 26, 959–1047.
- Blau, F., & Kahn, L. (2007). The gender pay gap: Have women gone as far as they can? *Academy of Management Perspectives*, 21, 7–23.
- Blumrosen, A. (1993). *Modern law: The law transmission system and equal employment opportunity*. Madison: University of Wisconsin Press.
- Boggs, R., Sellers, J., & Bendick, M., Jr. (1993). Use of testing in civil rights enforcement. In M. Fix and R. Struyk (Eds.), *Clear and convincing evidence: Measurement of discrimination in America* (pp. 345–376). Washington, DC: Urban Institute Press.
- Bolger, N., Davis, A., & Rafielli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616. doi: 10.1146/54.101601.145030.
- Bovenkerk, F. (1992). *Testing discrimination in natural experiments: A manual for international comparative research on discrimination on the grounds of "race" and ethnic origin*. Geneva: International Labour Office.
- Brewer, M. (1979). In-group bias in the minimal intergroup situation: A cognitive motivational analysis. *Psychological Bulletin*, 86, 307–324. doi: 10.1037/0033-2909.86.2.307.
- Brief, A. (Ed.) (2008). *Diversity at work*. New York: Cambridge University Press.
- Chartrand, T., & Bargh, J. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893–910. doi: 10.1037/0022-3514.76.6.893.
- Cialdini, R. (2000). *Influence: Science and practice*. Boston: Allyn & Bacon.
- Clark, A., Georgellis, Y., & Sanfey, P. (1999). Scarring: The psychological effect of past unemployment. *Economica*, 68(2), 221–241.
- Darity, W., & Mason, P. (1998). Evidence on discrimination in employment: Codes of color, codes of gender. *Journal of Economic Perspectives*, 12, 63–90.
- Darley, J., & Gross, P. (1983). A hypothesis-conforming bias in labelling effects. *Journal of Personality and Social Psychology*, 44, 20–33. doi: 10.1037/0022-3514.44.1.20.
- Dipboye, R., & Colella, A. (2005). *Discrimination at work: The Psychological and Organizational Bases*. Mahwah, NJ: Lawrence Erlbaum.
- Discrimination Research Center (2004). *Names make a difference: The screening of resumes by temporary employment agencies in California*. Berkeley, CA: Discrimination Research Center of the Impact Fund.
- Dovidio, J., & Gaertner, S. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11(4), 315–319. doi: 0.1111/1467-9280.00262.
- Dovidio, J., Kawakami, K., & Gaertner, S. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62–68. doi: 10.1037/0022-3514.82.1.62.
- Eagly, A., & Karau, S. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573–598. doi: 0.1037/0033-295X.109.3.573.
- Eberhardt, J., Dasgupta, N., & Banaszynski, T. (2003). Believing is seeing: The effects of racial labels and implicit beliefs on face perception. *Personality and Social Psychology Bulletin*, 29, 360–370. doi: 10.1177/0146167202250215.
- Edelman, L., & Petterson, S. (1999). Symbols and substance in organizational response to civil rights law. *Research in Social Stratification and Mobility*, 17, 107–136.
- Egan, M., & Bendick, M., Jr. (2008). Combining multicultural management and diversity into one course on cultural competence. *Academy of Management Learning and Education*, 7, 387–393.
- Egan, M., Bendick, M., Jr., & Miller, J. (2002). US firms' evaluation of employee credentials in international business. *International Journal of Human Resource Management*, 13, 78–88.

- Egan, M., Mauleon, F., Wolff, D., & Bendick, M., Jr. (2009). France's Mandatory "Triple Bottom Line" Reporting: Promoting sustainable development through informational regulation. *International Journal of Environmental, Cultural, Economic, and Social Sustainability*, 7, 27–47.
- Erber, R., & Fiske, S. (1984). Outcome dependency and attention to inconsistent information. *Journal of Personality and Social Psychology*, 47, 709–726. doi: 10.1037/0022-3514.47.4.709.
- Esen, E. (2005). *2005 workplace diversity practices: Survey report*. Alexandria, VA: Society for Human Resource Management.
- European Commission (2006). *Equality and non-discrimination annual report*. Brussels: Commission of the European Union.
- Finkelstein, L., Burke, M., & Raju, N. (1995). Age discrimination in simulated employment contexts: An integrative analysis. *Journal of Applied Psychology*, 80, 625–663. doi: 10.1037/0021-9010.80.6.652.
- Fiske, S., Bersoff, D., Borgida, E., Deaux, K., & Heilman, M. (1991). Social science research on trial: The use of sex stereotyping research in Price Waterhouse v. Hopkins. *American Psychologist*, 46, 1049–1060. doi: 10.1037/0003-066X.46.10.1049.
- Fiske, S., Cuddy, A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82, 878–902. doi: 10.1037/0022-3514.82.6.878.
- Fix, M., & Turner, M. (Eds.) (1999). *A national report card on discrimination: The role of testing*. Washington, DC: Urban Institute Press.
- Fogelman, E. (1994). *Conscience and courage, rescuers of Jews during the Holocaust*. New York: Anchor Doubleday.
- Freshman, C. (1990). Beyond atomized discrimination: Use of acts of discrimination against "other" minorities to prove discriminatory motivation under federal employment law. *Stanford Law Review*, 43, 241–273.
- Gaertner, S., & Dovidio, J. (2005). Understanding and addressing contemporary racism: From aversive racism to the common ingroup identity model. *Journal of Social Issues*, 61, 615–639. doi: 10.1111/j.1540-4560.2005.00424.x.
- Gallup (2005). *Employee discrimination in the workplace*. Washington, DC: The Gallup Organization.
- Gawronski, B., Deutsch, R., Mbirkou, S., Siebt, B., & Strack, F. (2008). When 'Just say no' is not enough: Affirmation vs. negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, 44, 370–377. doi: 10.1016/j.jesp.2006.12.004.
- Giacalone, R., & Rosenfeld, P. (1989). *Impression management in the organization*. Hillsdale, NJ: Lawrence Erlbaum.
- Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology*, 55, 178–186. doi: 10.1037/0022-3514.55.2.178.
- Golden, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90, 715–741.
- Goldsmith, A., Sedo, S., Darity, W., & Hamilton, D. (2004). The labor supply consequences of perceptions of employer discrimination during search and on-the-job: Integrating neoclassical theory and cognitive dissonance. *Journal of Economic Psychology*, 25, 15–39. doi: 10.1016/S0167-4870(02)00210-6.
- Granovetter, M. (1995). *Getting a job, A study of contacts and career*. Chicago: University of Chicago Press.
- Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464.
- Guardian (2006). *Good jobs NOW! A snapshot of the economic health of San Francisco's transgender communities*. San Francisco, CA: The San Francisco Bay Guardian.
- Heckman, J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, 12, 101–116.
- Hertel, G. & Kerr, N. L. (2001). Priming in-group favoritism: The impact of normative scripts in the minimal group paradigm. *Journal of Experimental Social Psychology*, 37, 316–324. doi: 10.1006/jesp.2000.1447.
- Higgins, C., & Judge, T. (2004). The effect of applicant influence tactics on recruiter perceptions of fit and hiring recommendations: A field study. *Journal of Applied Psychology*, 89, 622–632. doi: 10.1037/0021-9010.89.4.622.

- Hilton, J., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47, 237–271. doi: 10.1146/annurev.psych.47.1.237.
- HRSA (2010). *The physician workforce: Projections and research into current issues affecting supply and demand*. Washington, DC: U.S. Department of Health and Human Services, Health Resources and Services Administration.
- Hulett, D., Bendick, M., Jr., Thomas, S., & Moccio, F. (2008). Enhancing women's inclusion in firefighting in the USA. *International Journal of Diversity in Organizations, Communities, and Nations*, 8, 1–24.
- Ibarra, H. (1995). Race, opportunity, and diversity of social circles in managerial network. *Academy of Management Journal*, 38, 673–703.
- ILO (1998). *Discrimination in access to employment on grounds of foreign origin: The case of Belgium*. International Migration Paper 23. Geneva: International Labour Organisation.
- Johnson, J., Jr., Bienenstock, E., & Stoloff, J. (1995). An empirical test of the cultural capital hypothesis. *Review of Black Political Economy*, 23, 7–27.
- Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Diversity management and the remediation of inequality. *American Sociological Review*, 71, 589–617.
- Kang, J. (2012). The missing quadrants of anti-discrimination: Going beyond the "Prejudice Polygraph". *Journal of Social Issues*, 68(2), 314–327. doi: 10.1111/j.1540-4560.2011.01750.x
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, 78, 871–888. doi: 10.1037/0022-3514.78.5.871.
- Kenney, G., & Wissoker, D. (1994). An analysis of the correlates of discrimination facing young Hispanic job-seekers. *American Economic Review*, 84, 674–83.
- Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K., Leonard, J., Levine, D., & Thomas, D. (2003). The effects of diversity on business performance: Report of the diversity research network. *Human Resource Management*, 42, 3–21.
- Koomen, W., & Dijker, A. J. (1998). Ingroup and outgroup stereotypes and selective processing. *European Journal of Social Psychology*, 27, 589–601. doi: 10.1002/(SICI)1099-0992(199709/10)27:5<589::AID-EJS P840>3.0.CO;2-Y.
- Kuhn, P., & Skuterud, M. (2000). Job search methods: Internet versus traditional. *Monthly Labor Review*, 123, 3–11.
- Le, H., Oh, I., Shaffer, J., & Schmidt, F. (2007). Implications of methodological advances for practices of personnel selection: How practitioners benefit from meta-analysis. *Academy of Management Perspectives*, 21, 6–15.
- Linn, M., Sandifer, R., & Stein, S. (1985). Effects of unemployment on mental and physical health. *American Journal of Public Health*, 75, 507–512.
- Lodder, L. (1994). *Employment testing for civil rights enforcement: An operations manual*. Chicago: Legal Assistance Foundation of Chicago.
- Maass, A., Salvi, D., Acuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, 57, 981–993. doi: 10.1037/0022-3514.57.6.981.
- Macrae, C., Bodenhausen, G., & Milne, A. (1998). Saying no to unwanted thoughts: Self-focus and the regulation of mental life. *Journal of Personality and Social Psychology*, 74, 578–590. doi: 10.1037/0022-3514.74.3.578.
- Make the Road (2010). *Transgender need not apply: Gender identity job discrimination in New York City's retail sector*. New York: Make the Road New York.
- McConahay, J., Hardee, B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution*, 25, 563–579.
- Monin, B., & Miller, D. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81, 33–43. doi: 10.1037/0022-3514.81.1.33.
- Moore, R. (1990). *The formation of a persecuting society: Power and deviance in Western Europe* (pp. 950–1250). Oxford: Blackwell.
- Moss, P., & Tilly, C. (2002). *Stories employers tell: Race, skill, and hiring in America*. New York: Russell Sage Foundation.
- Murray, R. (1988). *Social closure: The theory of monopolization and exclusion*. Oxford: Oxford University Press.

- Nelson, T., Acker, M., & Manis, M. (1996). Irrepressible stereotypes. *Journal of Experimental Social Psychology, 32*, 13–38. doi: 10.1006/jesp.1996.0002.
- Nosek, B., Greenwald, A., & Banaji, M. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality & Social Psychology Bulletin, 31*, 166–180. doi: 10.1177/0146167204271418.
- Nunes, A., & Seligman, B. (1999). *Treatment of Caucasian and African-American applicants by San Francisco Bay Area employment agencies: Results of a study utilizing "testers."* Berkeley, CA: Discrimination Research Center of the Impact Fund.
- Nunes, A., & Seligman, B. (2000). *A study of the treatment of female and male applicants by San Francisco Bay Area auto service shops.* Berkeley, CA: Discrimination Research Center of the Impact Fund.
- Pager, D. (2007). The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *Annals of the American Academy of Political and Social Science, 609*, 104–133.
- Pager, D., & Western, B. (2012). Identifying discrimination at work: The use of field experiments. *Journal of Social Issues, 68*(2), 221–237. doi: 10.1111/j.1540-4560.2011.01746.x
- Peterson, B., Doty, R., & Winter, D. (1993). Authoritarianism and attitudes toward contemporary social issues. *Personality and Social Psychology Bulletin, 19*, 174–184. doi: 10.1177/0146167293192006
- Reskin, B., & Bielby, D. (2005). A sociological perspective on gender and career outcomes. *Journal of Economic Perspectives, 19*, 71–86.
- Riach, P., & Rich, J. (2002). Field experiments in discrimination in the market place. *The Economic Journal, 112*, F480–F518.
- Roberson, L., & Kulik, C. (2007). Stereotype threat at work. *Academy of Management Perspectives, 21*, 24–40.
- Rodgers, W., III (Ed.) (2006). *Handbook on the economics of discrimination.* Northampton, MA: Edward Elgar.
- Rooth, D. (2007). *Implicit discrimination in hiring: Real world evidence.* Discussion Paper 2764. Berlin: IZA.
- Rooth, D. (2009). Obesity, attractiveness, and differential treatment in hiring: A field experiment. *Journal of Human Resources, 44*, 710–735.
- Schilt, K., & Wiswall, M. (2008). Before and after: Gender transitions, human capital, and workplace experiences. *The Bell Economic Journal of Economic Analysis and Policy, 8*, 1–26.
- Semin, G., & Fiedler, F. (Eds.) (1992). *Language, interaction, and social cognition.* Newberry Park, CA: Sage Publications.
- Shelly, R., & Shelly, A. (2009). Speech content and the emergence of inequality in task groups. *Journal of Social Issues, 65*, 307–333. doi: 10.1111/j.1540-4560.2009.01602.x.
- Shelton, J., Richeson, J., & Salvatore, J. (2005). Expecting to be the target of prejudice: Implications for interethnic interactions. *Personality and Social Psychology Bulletin, 31*, 1189–1202. doi: 10.1177/0146167205274894.
- Sidanius, J., & Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression.* Cambridge: Cambridge University Press.
- Smith, J., & Welch, F. (1989). Black economic progress after Myrdal. *Journal of Economic Literature, 27*, 519–564.
- Smith, T. (2000). *Taking America's Pulse II, NCCJ's 200 survey of intergroup relations in the United States.* New York: National Conference for Community and Justice.
- Stangor, C., Sechrist, G., & Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Personality and Social Psychology Bulletin, 27*, 486–496. doi: 10.1177/0146167201274009.
- Steele, C. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.) *Advances in experimental social psychology* (Vol. 21, pp. 261–302). Orlando, FL: Academic Press.
- Steele, C., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811. doi: 10.1037/0022-3514.69.5.797.

- Stockdale, M., & Crosby, F. (2004). *The psychology and management of workplace diversity*. Malden, MA: Blackwell.
- Swim, J., & Sanna, L. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin*, *22*, 507–519. doi: 10.1177/0146167296225008.
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*, 1–39.
- Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorisation and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178. doi: 10.1002/ejsp.2420010202.
- Tal, A., Moran, G., Rooth, D., & Bendick, M., Jr. (2009). Using situation testing to document employment discrimination against persons with psychiatric disabilities. *Employee Relations Law Journal*, *35*, 82–102.
- Tomaskovic-Devey, D. (1993). The gender and race composition of jobs and the male/female, white/black pay gaps. *Social Forces*, *72*, 45–76.
- Townsend, B. (1997). *Confidential study of performance evaluations*. New York: Catalyst.
- Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse and Society*, *14*, 191–220.
- U.S. Bureau of Labor Statistics (2008). Employed persons by detailed occupations, sex, race, and hispanic or latino ethnicity (Downloaded February 10, 2010 from www.bls.gov/cps/cpsaat11.pdf).
- U.S. Bureau of Labor Statistics (2009). Job openings and labor turnover-August 2009 (Downloaded October 26, 2009 from www.bls.gov/news.release/pdf/jolts.pdf).
- U.S. Equal Employment Opportunity Commission (EEOC) (2010). Charge statistics FY 2007 through FY 2009 (Downloaded February 22, 2010 from www.eeoc.gov/eeoc/statistics/enforcement/charges.cfm).
- Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: MIT Press.
- Word, C., Zanna, M., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interactions. *Journal of Experimental Social Psychology*, *10*, 109–120. doi: 10.1016/0022-1031(74)90059-6.
- Yelnosky, M. (2010). *Testers Revisited*. Legal Studies Paper 74. Bristol, RI: Roger Williams School of Law.
- Zitek, E. M., & Hebl, M. R. (2007). The role of social norm clarity in the influenced expression of prejudice over time. *Journal of Experimental Social Psychology*, *43*, 867–876. doi: 10.1016/j.jesp.2006.10.010.

MARC BENDICK, JR., is a Principal in Bendick and Egan Economic Consultants, Inc., in Washington, DC. He received his PhD from the University of Wisconsin. His 125 scholarly publications concern poverty, employment, and public policies to enhance inclusion of individuals, businesses, and communities in the economic mainstream. He is a consultant on workforce diversity management to major employers, frequent expert witness in discrimination litigation, and has led dozens of matched pair testing studies.

ANA P. NUNES is a post-doctoral social psychology fellow at the Instituto de Ciências Sociais da Universidade de Lisboa (Institute of Social Sciences of the University of Lisbon). She received her Doctorate in social psychology from the University of Colorado Boulder. Her research interests focus on person perception processes and the effect of diversity mandates on judgments and behavior, especially in high-stake decision contexts such as hiring. She formerly directed testing studies of discrimination in housing for Project Sentinel in Palo Alto, CA, and employment for the Discrimination Research Center in Berkeley, CA.