Running Head: BIAS IN HIRING

February 27, 2010

This paper is an invited contribution to the issue

"The Reality of Contemporary Discrimination,"

forthcoming in the *Journal of Social Issues*, 2011

DEVELOPING THE RESEARCH BASIS FOR CONTROLLING BIAS IN HIRING

MARC BENDICK, JR.[*]

BENDICK AND EGAN ECONOMIC CONSULTANTS, INC., WASHINGTON, DC

ANA P. NUNES

DEPARTMENT OF PSYCHOLOGY AND NEUROSCIENCE, UNIVERSITY OF COLORADO AT BOULDER

Abstract

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

DEVELOPING THE RESEARCH BASIS FOR CONTROLLING BIAS IN HIRING

Employers in America's private sector decide who to hire about 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 also determines initial job titles, work assignments, and wages, which in turn often influence that employee's career for years thereafter. For both 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 terms of impact on the distribution of employment opportunities and rewards.

Controlling bias that potentially distorts these high-stakes decisions has long been a goal of American society. This concern is reflected in federal, state, and local laws against hiring discrimination, as well as legal provisions for affirmative action in recruitment and hiring. 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).

As Section II of this paper reviews, research in psychology, sociology, economics and management offers many insights into the mechanisms of bias and the likely effects of interventions to mitigate their effects. However, as Sections III and IV document, despite considerable progress, bias continues to pervade the American hiring system. Accordingly, the paper proposes additional research to support further changes in employer hiring practices. Section V proposes four types of studies to expand conceptual understanding of hiring bias, while Section VI proposes three types of studies to enhance the impact of such research on the behavior of hiring practitioners.

A number of other publications synthesize research on employment bias and identify future research needs (e.g., Brief, 2008; Dipboye & Colella, 2005; Stockdale & Crosby, 2004). The present paper differs from those reviews by focusing on hiring and one innovative research methodology — matched pair testing — with unique potential for studying hiring. For readers not familiar with matched pair testing, it is briefly described in an appendix.

II. 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 demographic groups. Correspondingly, ensuing anti-discrimination legislation — prominently, the Civil Rights Act of 1964 (mandating equal opportunity based on race, gender and selected other personal characteristics), 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 to address 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. Equally, it suggests the need to improve specific employment "micro 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 either explicitly or implicitly.

Stereotypes Infect Us All

The first such line of research explores stereotypes and their role in discriminatory behavior. *Implicit bias* — unconscious association of specific traits with members of a demographic group — has been extensively measured with the Implicit Association Test (IAT) (Greenwald, McGhee, and Schwartz, 1998; www.implicit.harvard.edu). Studies employing the IAT use experimental subjects' reaction times to illustrate that, regardless of their expressed or self-perceived attitudes toward racial equality, individuals tend to associate negative traits with racial minorities (Nosek, Banaji, & Greenwald, 2002; Nosek, Greenwald, & Banaji, 2005). Those implicit associations, in turn, correlate with biased behavior (Dovidio, Kawakami, & Gaertner, 2002). When such associations are activated in the hiring process, the predominantly negative content of stereotypes about traditionally-excluded groups (e.g. African Americans are uneducated; women are not career-committed), tends to handicap members of these groups in competing for jobs.

Decades of related research has further elucidated the processes by which stereotypes unconsciously influence how individuals are perceived and evaluated, processes 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 & Dijker, 1998), interpret ambiguous information to confirm stereotypes (Hilton & Van Hippel, 1996), seek out 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 scores on skill tests (Darley &

Gross, 1980). Together, such processes explain how hiring decision-makers may honestly

perceive themselves as making unbiased selections that reflect 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 individual members of

stigmatized groups face when they attempt to ameliorate the adverse effects of negative

stereotypes about the groups to which they belong.   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 from others than if they had conformed

to the stereotype.

Compounding these difficulties, biased treatment itself or anticipation of it, can adversely

affect the actual performance of employees or job applicants. For example, Word, Zamma, and

Cooper (1974; see also Shelton, Richeson, & Salvatore, 2005; Richeson, Trawalter, & Shelton,

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

their interviews, and that this interviewer behavior caused applicants to perform less well during

interviews.

Deficits in minority applicants' performance during job interviews may also be triggered

by *stereotype threat.* Stereotype threat research (Steele & Aronson, 1995; Roberson & Kulik,

2007) suggests that minorities and women tend 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). A plausible explanation for such effects is that stereotype threat enhances anxiety (O'Brien & Crandall, 2003).

Shared Traits Confer Advantage

A third line of research focuses on *ingroup bias*, or the tendency for individuals to prefer members of their own ingroup and derogate those of other groups, as a function of the need to maintain a positive social identity (Brewer, 1979; Tajfel, 1982). Studies have documented that this unconscious process may lead to employment decisions not based on applicants' abilities to perform on the job, but rather the decision-maker's categorization of the applicant as an ingroup member. This ingroup preference has been empirically demonstrated in the laboratory even with minimal groups, a fabricated group to which an experimental subject is lead to identify as a member   (Tajfel, Billig, Bundy, & Flament, 1971; Hertel & Kerr, 2001). 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, & Raju (1995) concluded that young persons with authority to hire tended to rate young job applicants more favorably than older ones.

Employment advantage conferred by shared traits may also be created by social patterns that extend beyond the workplace. Many social relationships in American society today remain highly segregated. Persons of different racial, ethnic, and class backgrounds tend to reside in separate neighborhoods and attend different schools and churches. Persons of different genders and ages tend to socialize in different groups and social networks (Alba, Logan, & Stutts, 2000). But the social relationships created through such interactions often are resources in the labor market (Granovetter, 1995; Adler & Kwon, 2002; Ibarra, 1995). For instance, many job

vacancies are never publicly advertised (e.g., on internet job boards or in newspapers), so that the opportunities become known primarily to the friends, relatives, neighbors, classmates, or colleagues of current employees (Kuhn & Skuterud, 2000). Similarly, in pre-employment skill testing and job 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, anti-discrimination legislation has certainly reduced many blatant forms of employment discrimination, such as the traditional division of job vacancy announcements between "Help Wanted-Male" and Help Wanted-Female." Over the same period, public opinion polls report consistent downward trends in individuals' self-reported prejudice. For example, in public opinion polling in 1965, 59% of Americans agreed that they would vote for a qualified African American for president, 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, research cautions that these developments may not translate into substantially reduced bias in employment. Rather, blatant discriminatory behavior toward minority groups may have simply evolved into what McConahay, Hardee, and Batts (1981) label *modern racism* -- a shift in social norms such that explicit expression of prejudiced attitudes is no longer tolerated and rarely seen, yet prejudiced attitudes remain in more covert forms.

Consistent with the modern racism model, researchers have observed such covert prejudice reflected in the behavior of actual supervisors and managers. For example, Moss and

Tilly (2002; see also Neckerman & Kirschenman, 1991, and Rooth, 2007) describe how, in rejecting minority job applicants, hiring decision-makers now often cite those applicants' deficiencies in "soft skills" (such as "inability to communicate" or "lack of work commitment") as a socially-acceptable proxy for a job applicant's race. Other researchers have described some employers' initiatives to advance equal employment opportunity as primarily symbolic acts, creating the appearance of compliance with the law and social norms without the intention to change employment outcomes substantially (Edelman & Petterson, 1999; Bendick & Egan, 2009).

Hiring is Particularly Vulnerable to Bias

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

The first of these circumstances is the limited information on which hiring decisions are based (Altonji & Pierret, 2001).   Post-hiring decisions tend to be made by managers and supervisors who have observed an employee's performance over months or years. Hiring decisions, in contrast, typically require assessing job applicants who are virtual strangers. Applicants' resumes, application forms, and work samples typically provide only a sketch of their qualifications and require inference about the relevance of those qualifications 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, 2009). Interviews also tend to be highly subject to "impression management," or deliberate manipulation of self-

presentation to create a favorable impression that does not necessarily translate to successful job performance (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 that negatively impacts members of stigmatized groups is the time pressure under which hiring decisions are often made. Supervisors and managers often must assess numerous job applications while facing other demands on their time. Stereotypes tend to exercise particular influence in time-pressured situations (Macrae, Bodenhausen, & Milne, 1999). Concurrently, limited contact between job applicants and hiring decision-makers imbues every interaction with intense pressure to perform making it particularly susceptible to stereotype threat.

A different circumstance enhancing bias in hiring compared to post-hiring decisions is differences in the opportunity to correct errors. Many post-hiring employment decisions are part of ongoing or repeated decision processes in which an error can be corrected later. For instance, an employee granted too generous a raise in one year may be brought back into line by a smaller raise the next year, or an employee passed over for one promotion may be subsequently considered for other opportunities. In fact, the continued visibility of a minority employee who is not progressing may itself pressure managers to correct the situation (Zitek & Hebl, 2006; Stangor, Sechrist, & Jost, 2001). In hiring, in contrast, applicants who are not hired tend immediately to go elsewhere to find employment, eliminating opportunities for the employer to whom they had applied to reconsider or adjust.

External pressure to correct bias, such as through enforcement of anti-discrimination laws, is also less likely for hiring decisions than post-hiring decisions. An employee who feels

aggrieved about a post-hiring decision often has both the information necessary to file a

complaint (e.g., Who was promoted instead of me?) and a sufficient stake in the outcome to

invest the resources need to seek redress. In contrast, a job applicant who suspects hiring

discrimination often lacks sufficient information to be sure that discrimination has occurred (e.g.,

When the company said the job vacancy was already filled, was that the truth or an excuse?), and

is more likely to pursue other job opportunities than to dispute the decision. For such reason,

hiring complaints comprise only 6% of formal discrimination complaints filed with the Equal

Employment Opportunity Commission (Bendick, Jackson, & Reinoso, 1994; see also Donohue

& Siegelman, 1991).

III. Empirical Evidence of Employment Bias

The research reviewed in Section II can be translated 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 indications of change reveal as much about remaining problems as they do

about progress. Earnings ratios of 73% or 80% are still far below the 100% which would signal

simple equality. Worse, in recent years, upward movement of these race and gender wage ratios has slowed and perhaps plateaued (Rodgers, 2006). Concurrently, the representation of women and minorities remains very limited in many occupations, especially more prestigious, well paid ones; these patterns of occupational segregation are 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. Women still account for only about 1% of skilled construction craft workers, such as carpenters and electricians (U.S. Bureau of Labor Statistics, 2008). And although women now account for 31% of medical doctors, they are only 9% of cardiologists and 4% of orthopedic surgeons. Moreover, female physicians earn an average of 18% less than male physicians with equal credentials, medical specialties, years in practice, and hours of work per week (HRSA, 2010).

Research in the form illustrated by these findings on physicians' earnings -- showing different employment outcomes among demographic groups that are 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, and localities for many different demographic groups, producing results such as the following:

- After controlling for their 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 & Wiwsall, 2008).

- Among American professionals and managers working in international business, an additional year of international work experience raised men's salaries an average of $2,500 but women's only $1,300; working an extra 8 hours a week increased men's salaries an average of $7,300 but women's only $4,300 (Egan, Bendick, & Miller, 2002).

- In the state of Georgia, 28% of large employers employed statistically-significantly fewer minority workers than peer companies in the same industry hiring workers with the same qualifications and local labor market (Blumrosen, Bendick, Miller, & Blumrosen, 1999; see also Bendick, 2000).

- Leading symphony orchestras which auditioned musicians behind screens so that judges could not see the job seekers' sex hired more women players than those which conducted auditions in the open (Golden & Rouse, 2000).

Such findings are consistent with less rigorous evidence based on perceptions of bias reported by adversely affected individuals. In one typical public opinion 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 still another national survey, 53% of workers of all backgrounds recalled hearing jokes at work about women, 44% for jokes about sexual orientation, 38% about older people, 37% about African

Americans, 32% about Hispanics, 29% about Jews, 29% about Muslims, 26% about people with

disabilities, and 25% about Asians (Harris Interactive, 2002). In narrower examples, in a survey

of 194 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, in forms ranging from hostile comments to chronic sexual

advances in firehouses and being placed in unnecessary danger on fire scenes (Hulett, Bendick,

Thomas, & Moccio, 2008).

Finally, evidence of perceived discrimination is provided by formal legal complaints of

employment discrimination filed with the U.S. Equal Employment Opportunity Commission,

which in 2008 numbered 93,000 (U.S. Equal Employment Opportunity Commission, 2010). In

parallel, thousands of anti-discrimination lawsuits, both by the public anti-discrimination

enforcement agencies and by private litigators, continue to be filed each year and settled in favor

of plaintiffs, with some settlements including tens of millions of dollars in compensation for

discrimination-based earnings losses (Darity & Mason, 1998; Bendick & Egan, 2009).

IV. 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 employment processes; may be based

on perceptions rather than objectively-verified facts; sometimes requires extrapolation to the

behavior of actual employment decision-makers from findings demonstrated in laboratory

settings; or is indirectly inferred from observing differences in employment outcomes that

remain after other job-related explanations have been eliminated. Since about 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 (see Appendix: A Primer on Matched Pair Tests of Hiring).

Over the past two decades, several dozen matched pair testing studies have examined hiring in the Unites States in labor markets from Boston to Los Angeles and occupations from entry-level retail sales to professional and managerial positions (Bendick, 2007; Pager, 2007; Bendick, 1999). Several dozen additional studies have been conducted in other industrialized nations, ranging from the Netherlands to Australia (Riach & Rich, 2002; European Commission, 2006; ILO, 2010). The 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 various European nations.

These studies have been essentially unanimous in documenting considerable bias in hiring. Among studies conducted in a rigorous manner which analyzed the complete hiring process, the estimated net rate of discrimination typically ranges between 20% and 40%. That is, the studies found that job applicants from historically excluded groups had a 20% to 40% chance of a substantially less favorable outcome because of that group membership each time they applied for a job. Equivalently, these studies estimate that bias infects the hiring decisions of 20% to 40% of employers.

Such estimates are striking consistent with the non-testing evidence reviewed in Sections II and III. In falling far below 100%, the estimate rates of discrimination reflect the progress toward employment equality over the past four decades; in falling far above 0%, they confirm that substantial bias remains.

One obvious direction for future matched pair testing studies is to continue to measure the extent of discrimination in labor markets, systematically "mapping" the prevalence of discrimination in different locations (Are workplaces in the suburbs more discriminatory than central cities?), occupations and industry (Are higher paid jobs more subject to bias than entry-level jobs?), and demographic groups (Do women or African Americans face worse problems?) Such studies would be particularly useful if they were repeated annually using a consistent methodology on a nationally-representative sample to generate a periodic "national report card" on discrimination tracing trends over time (Fix & Turner, 1999).

A related form of "mapping" has been conducted in recent years through innovative studies that have applied testing to individuals facing bias on bases other than the "usual" race, gender, age, and disability characteristics. These studies have involved, for instance, Arab Americans facing backlash following the terrorist attacks of September 11, 2001 (Discrimination Research Center, 2004), persons with a criminal record (Pager, 2003), transgender individuals (Make the Road New York, 2010), and overweight persons (Rooth, 2009).

These testing studies have typically revealed relatively modest variation in the prevalence of discrimination among demographic groups, locations, and occupations, suggesting that continued testing to explore how rates of discrimination vary across demographic groups or labor market circumstances should perhaps no longer 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 ballot proposition to abolish affirmative action (Proposition 209) (Bendick, 1995), the proposition passed.

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

V. 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 bias that occurs at the implicit level, is often embodied in small, subtle but crucial differences in words or actions sometimes referred to as *micro-inequities* (Valian, 1998). These small, subtle differences often leave perpetrators of bias unaware of these unconscious processes and their cumulative impact. Equally, they pose measurement challenges to researchers that seek to document exactly how bias operates.

To date, research on the details of employer-employee interactions has often examined written materials such as letters of recommendations and performance evaluations (e.g., Trix & Psenka, 2003; Bison-Rapp, 1999). These studies have documented striking systematic differences in statements about equally-qualified individuals of different demographic backgrounds. For example, one comparison of performance evaluations for men and women

professional employees who all received high performance ratings in the same financial services firm found that women were praised primarily for activities within their own work group and men for activities involving external relationships; 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 ingroup and outgroup behavior that itself tends to perpetuate stereotypes — for instance, by describing outgroup positive behavior and ingroup negative behavior in concrete terms that suggest situational and therefore not enduring traits, while reporting outgroup negative behavior and positive ingroup behavior in abstract terms that suggest persistent qualities (Maass, Salvi, Arcuri, & Semin, 1989; see also Semin & Fiedler, 1992, and Wenneker, Wigboldus, & Spears, 2005).

Only a few studies in this tradition have examined face-to-face interactions rather than written material (Binning, Goldstein, Garica, & Scatteregia, 1988, and Shelley & Shelley, 2009. The data required for such research are samples of the exact words used to describe demographically different individuals who are equivalent in actual qualifications or actual performance. Filling a gap in existing research, matched pair testing could provide such data for in-person interactions, such as job interviews. Through modern miniature recording technology, voice and video recordings of conversations can now be collected while leaving employers unaware of being tested or recorded.   Transcripts of these encounters could then be analyzed using the psycholinguistic techniques and concepts illustrated in the previous paragraph.

Some testing studies have examined employer and employee behavior in job interviews. For example, Bendick, Jackson, & Reinoso (1994) examined interviews of African American

and White applicants for entry-level positions that do not require college degrees. They

observed that, although the two groups received very different numbers of job offers, applicants

experienced no substantial differences in treatment while being interviewed. Here, applicant

treatment was measured using observable behavior such as interview length, proportion of the

interview devoted to job-relevant topics rather than social chit chat, 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, 2009). It documented that

interviewers tended to accept past restaurant experience without probing when it was claimed by

White applicants but posed aggressive, skeptical questions to non-White applicants claiming

equivalent experience. Such studies are far from definitive, but they suggest the richness of

insights that more sophisticated analyses of interview transcripts could generate.

Coordinate Testing Studies of Multiple Outgroups

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

the same employer.

Many analyses of discrimination against historically excluded groups focus on issues

specific to each group, for example, the impact of slavery and 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

the same 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

outgroups but rather *social closure* to preserve the power, status, and privilege of a dominant ingroup against *all* alternative claimants (Tomascovic-Devey, 1993; Murray, 1988; Moore, 1990; Freshman, 1990). In this model, an employer that discriminates against women will tend to do the same against race-ethnic minorities because both actions preserve White male dominance.

This debate is important not only to social theory but also to practical anti-discrimination efforts. For instance, the first perspective implies that it is important to contradict the stereotypes associated with individual outgroups (e.g., older workers are obsolete; women are not good leaders). In contrast, the second perspective considers stereotypes primarily *ex post* rationales for negative treatment of outgroups rather than independent causes and suggests controlling the generic process of stereotyping rather than addressing the content of specific stereotypes (Bendick, Egan, Lofhjelm, 2001; Egan & Bendick, 2008).

To date, matched pair testing has typically examined one outgroup per study. However, it could be organized to, for example, field teams of males and females, African Americans and Whites, and older and younger workers to apply for jobs with the same sample of employers. Positive correlations in bias against multiple outgroups would support the social closure model, while zero or negative correlations would support the alternative and imply that stereotype content does matter (Fiske, Cuddy, Glick, & Xu, 2002).

Behavioral science research also suggests other ways in which an employer's behavior toward an outgroup may differ depending on whether that group is considered in isolation or in a broader context that includes other outgroups. For instance, *moral credentialing* (Monin & Miller, 2001) suggests that an employer who has hired an applicant from a stigmatized group (e.g., women) and therefore feels that he or she has adequately demonstrated egalitarian values, may be less likely to evaluate favorably a subsequent female or minority candidate. Again,

testing studies could be designed to test this hypothesis, in this case, by appropriately sequencing

multiple tests of the same employer and then analyzing the relationship between each test

outcome and that of preceding tests.

Evaluate Strategies for Minimizing Bias

Practical advice routinely offered to outgroup job seekers includes a variety of

suggestions concerning how to minimize bias they may encounter.   For example, with respect to

stereotypes, some advisors counsel explicitly refuting stereotypes about groups to which the job

seeker belongs (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000); others suggest refuting

using individuating information to counter the applicability of the stereotype to the job seeker

(Beckett & Park, 2005; 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 these strategies proactively — in

anticipation that stereotypes will inevitably bias hiring decision-makers — while others suggest

doing so only if some indication triggers a concern.

Matched pair testing could be designed to evaluate such alternative strategies in actual

hiring situations.   Such research would provide advice to job seekers based on empirical

evidence in place of the current basis, which is largely intuitive.

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, the older applicant described

herself/himself as career-committed, energetic, and technologically up-to-date — positive

attributes stereotypically associated with younger workers; an alternative letter described the

older applicant 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 to the older applicant than the other

two (Bendick, Jackson, and Romero, 1996).

Employers also have implemented a variety of initiatives designed to minimize bias, and

again the empirical evidence supporting these practices is often intuitive rather than empirical.

In particular, most major employers in the United States today invest in "workforce diversity

training" for their employees, despite research which questions the effectiveness of many of

these efforts (Kalev, Dobbin, & Kelly, 2006; Bendick, Egan & Lofhjelm, 2001).

Matched pair testing studies could be structured to provide more credible assessments of

these efforts. For example, matched pair tests could be conducted in a workplace prior to a

diversity training program and then again after the training. Or a firm such as a large retail chain

could implement training in one subset of its establishments and not in a matched subset and then

conduct hiring tests in both the experimental and control locations. Here, testing could measure

not only the effect of the training on overall rates of discrimination but also on specific hiring

practices and behaviors that employees were trained to adopt or avoid.

Testing could also improve guidance for employers on other complex issues in bias

control. For example, in sincere efforts to avoid violating social norms, some hiring interviewers

become hyper-vigilant and consciously avoid any reference to race. However, Apfelbaum,

Sommers, & Norton (2008) demonstrate that such *strategic colorblindness* may backfire,

leading to nonverbal unfriendliness interpreted by Black observers as negative and prejudiced.

Similarly, Chartrand & 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 work performance. Matched pair 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. Findings could then be translated into practical guidance to interviewers in interviewing and assessing job candidates.

Study the Effects of Hiring Rejection

Behavioral science 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 in seeking work or investing in training (e.g., Linn, Sandifer, & Stein, 1985; Clark, Georgellis, & Sanfey, 1999). A few studies have focused on these effects when bias was a factor (Goldsmith, Sedo, Darity, & Hamilton, 2004). These studies employ a range of research techniques, including epidemiological studies of morbidity and mortality, worker surveys utilizing scales of self-efficacy and psychological well being, laboratory studies tracking physiological responses to disappointment, or diaries tracking attitudes and perceptions over time (Bolger, Davis, & Rafieli, 2003).

Matched pair testing offers a number of advantages as an additional methodology for measuring these effects. It provides unusually detailed information about the experimental stimulus whose effects are being studied -- exactly what jobs were sought, what testers experienced, and the bias encountered. It allows examining hiring separately from other aspects of workers' unemployment experiences, such as job separation. 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 workers with more experience.

Using such measures, researchers might study for example, the cumulative effects of multiple job disappointments; whether the effects of failing to be hired are different when the

decision was biased or unbiased; and whether the same disappointing hiring 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 on their self-esteem from failure to be hired than comparable 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 achieving success in a different domain (e.g., volunteer work).

In most testing studies, testers are research assistants who did not actually want the jobs to which they apply, and therefore their reactions to their testing experiences may differ substantially from those of actual job seekers. Accordingly, for research such as is suggested here, real job seekers should be hired as testers and dispatched to apply for jobs they would be allowed to take if offered.

However, testing using conventional research assistants can be used to study the effect on individuals' attitudes and empathy of being personally exposed to the reality of discrimination which they otherwise might not encounter. Research by Batson, *et al*. (2003; see also Gaertner & Dovisio, 2005, and www.janeelliott.com) concludes that putting an individual "in the shoes of others" expands the range of persons about whose well-being they feel concerned; Fogelman (1994) has documented the role of striking personal incidents in persuading bystanders to act on behalf of victims of persecution. In the same spirit, it can be hypothesized that ingroup testers (e.g. Whites or males) who observe their paired testing partners encounter bias may be made

more empathetic by the experience. This hypothesis could be studied by measuring the attitudes

and behavior of ingroup testers before and after their testing experience.

<div align="center">

VI.   Using Testing to Change Employer Behavior

</div>

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

lessons from research discussed throughout this paper into their firm's employment practices and

the behavior of their employees?   Matched pair testing offers new approaches for doing so, of

which this section sketches three proposals.

Testing on Behalf of Employers

Employers commonly monitor the behavior of their own staff using unobtrusive data

gathering techniques. For example, retailers often employ "mystery shoppers" to pose as

customers and record detailed data on their interactions with sales staff (www.mysteryshop.org).

They then use these data to improve the quality of service provided to their customers, such as

through staff retraining.

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 or for 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 its hiring practices were not biased. To support this

goal, the hospital required all its supervisors to be trained to conduct job interviews in a job-

related, structured manner (Le, Oh, Shaffer, & Schmidt, 2007).   However, the hospital had never

measured the extent to which these supervisors subsequently followed those procedures when interviewing job candidates or whether their hiring decisions resulted in a more diverse workforce. Accordingly, the hospital's vice president of human resources secretly arranged with a local non-profit 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. Although this project was not completed due to a hiring freeze, it exemplifies a potentially fruitful, cooperative relationship between testing researchers and an employer.

Businesses currently invest huge numbers of staff hours as well as many millions of dollars each year in initiatives 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 on the extent to which initiatives effectively modify employment practices "on the shop floor" and the extent to which those modifications make a difference in hiring outcomes.

Testing for Training

When testing results are presented to audiences of non-researchers — for example, in legislative hearings or articles in the mass media — their attention-grabbing power is immediately evident.   This power appears to derive from 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 taken advantage of in practical employee training. Among larger employers, more than 90% provide some form of anti-discrimination or

pro-diversity training for employees, often enrolling everyone from senior executives and mid-level supervisors to non -managerial employees (Esen, 2005; Kalev, Dobbin, & Kelly, 2006). The most effective forms of this training use vivid examples from real workplaces to persuade trainees of the continued presence of discrimination and to communicate desired changes in employee behavior (Adamson, 2000; Bendick, Egan, & Lofhjelm, 2001).

Testing results are ideally structured to provide specific, vivid examples for use in training, especially if tests are documented through audio or video recording as was suggested in Section V. As Section V also suggested, post-training testing of hiring outcomes could be used to assess whether training that includes testing-based illustrations is more effective than other approaches.

Testing for Litigation

Employers in the United States who violate the equal employment provisions of federal or state laws can be sued either by government agencies, such as the federal Equal Employment Opportunity Commission, 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 its 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, 1999).

American law gives matched pair testing potential roles in enforcement. If an employer is sued based on evidence other than testing, documentation of an employer's discriminatory behavior through testing can be used to corroborate that evidence. In addition, individual testers and non-profit organizations employing testers have "standing" to become plaintiffs in litigation based on testing evidence alone (Yelnosky, 2010; Boggs, Sellers, & Bendick, 1993).

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 non-profit organization that had hired them as testers. The defendant was a local office of Snelling & Snelling, one of the nation's largest job placement agencies. The agency had interviewed, coached, and found entry-level office jobs for two testers while it failed to do so for the African Americans who were their testing partners. This litigation was subsequently settled with payment of damages to the non-profit organization and commitments by the defendant to retrain its staff and implement other actions to eliminate discriminatory practices (Boggs, Sellers, & Bendick, 1993).

That lawsuit happened to have 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 anti-discrimination, 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 is 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 matched pair testing may be the only feasible way to document hiring discrimination in a manner that will hold up in court.

Litigation is only one adversarial approach in which testing could be used to pressure employers. To date, publicly-released results from testing studies have typically not named

employers against whom incriminating findings were obtained. However, the results of situation

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

pressure discriminating employers to change their behavior (Egan, Mauleon, Wolff, & Bendick,

2007).

VII. Summary

The seven types of matched pair testing studies discussed here form an ambitious

research agenda. However, these examples by no means exhaust the potential of matched pair

testing to generate theoretical and practical advances against hiring bias. 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,* 17-40.

Alba, R., Logan, J., & Stutts, B. (2000). How segregated are middle-class African Americans? *Social Problems*, 47, 543-558.

Altonji, J., & Blank, R. (1999). Race and gender in the labor market. In O. Aschenfelter & Card, D. (Eds.) *Handbook of Labor Economics*, *Volume 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.

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.

Beckett, N., & Park, B. (1995). Use of category versus individuating information: Making base rates salient. *Personality and Social Psychology Bulletin, 21*, 21-31.

Bell, M. (2007). *Diversity in Organizations*. Mason, OH: Thomson South-Western.

Bendick, Jr., M. (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: The Urban Institute.

Bendick, Jr., M. (2008) Measure Inclusion, Not Diversity. *Presentation, Society for Human Resource Management National Diversity Conference*.

Bendick, Jr., M. (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, Jr., M. (2007). Situation testing for employment discrimination in the United States of America. *Horizon Strategiques*, 5, 17-39.

Bendick, Jr., M. (2000). Using EEO-1 data to analyze allegations of employment discrimination. *Presentation, American Bar Association National Conference*.

Bendick, Jr., M., and 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, Jr., M., Egan, M., & Lofhjelm, S. (2001). Diversity training: From anti-discrimination compliance to organization development. *Human Resource Planning, 24,* 10-25.

Bendick, Marc, 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, Jr., M., Jackson, C., & Reinoso, V. (1994). Measuring employment discrimination through controlled experiments. *Review of Black Political Economy, 23*, 25-48.

Bendick, Jr., Marc, Jr., Jackson, C., Reinoso, V., & Hodges, L. (1991). Discrimination against Latino job applicants: A controlled experiment. *Human Resource Management, 30,* 469-484.

Bendick, Jr., M., 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, Jr., M., Rodriguez, R., & Jayaraman, S. (2009). *Race-ethnic employment discrimination in upscale restaurants: Evidence from matched pair testing*. Washington, DC: Bendick and Egan Economic Consultants, Inc, for the Restaurant Opportunity Center of New York.

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.

Binning, J., Goldstein, M., Garcia, M., & Scattereregia, J. (1988). The effects of preinterview impressions on interviewer questioning strategies in same- and opposite-sex employment interviews. *Journal of Applied Psychology, 73*, 30-37.

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.

Blumrosen, A., Bendick, Jr., M., Miller, J., & Blumrosen, R. (1999). *Employment discrimination against women and minorities in Georgia*. Rutgers, NJ: Rutgers University Law School.

Boggs, R., Sellers, J., and Bendick, Jr., M. (1993). Use of testing in civil rights enforcement. In Fix, M. and Struyk, R. (Eds.), *Clear and convincing evidence: Measurement of discrimination in America* (pp. 345-376). Washington: Urban Institute Press.

Bolger, N., Davis, A., & Rafieli, E. (2003). Diary methods, Capturing life as it is lived. *Annual Review of Psychology, 54,* 579-616.

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.

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.

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*, 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.

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.

Donohue, J., & Siegelman, P. (1991). The changing nature of employment discrimination litigation, *Stanford Law Review, 43, 983-1033*.

Dovidio, J., Kawakami, K., & Gaertner, S. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62-68.

Eagley, A., & Karau, S. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*, 573-598.

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.

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, Jr., M. (2008). Combining multicultural management and diversity into one course on cultural competence. *Academy of Management Learning and Education, 7*, 387-393.

Egan, M., Bendick, Jr., M., & 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, Jr. M. (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.

Esen, E. (2005). *2005 workplace diversity practices: Survey report.* Alexandria, VA: Society for

    Human Resource Management.

European Commision (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.

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.

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.

Fix, M. and Turner, M. (Eds.) (1999). *A national report card on discrimination: The role of testing*. Washington: 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.

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.

Giacalone, R., & Rosenfeld, P. *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.

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.

Granovetter, M. (1995). *Getting a job, A study of contacts and career*. Chicago: University of Chicago Press.

Greenwald, A., McGeer, D., and Schwartz, J. (1998). Measuring individual differences in implicit cognition, The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464 –1480.

Guardian (2006). *Good jobs NOW! A snapshot of the economic health of San Francisco's transgender communities*. San Francisco: The San Francisco Bay Guardian.

Harris Interactive (2002). *Workplace discrimination against, and jokes about, African Americans, gays, Jews, Muslims, and other*. The Harris Poll 61, at http://www.harrisinteractive.com, downloaded February 2, 2010.

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.

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.

Hilton, J., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47*, p. 237-271.

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, Jr., M., 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 (2010). *Discrimination: The evidence*, at www.ilo.org/public/english/protection/migrant/equality/discrimination/evidence.htm, downloaded February 17, 2010.

Johnson, Jr., J., 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., and Kelly, E. (2006). Best practices or best guesses? Diversity management and the remediation of inequality. *American Sociological Review, 71,* 589-617.

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.

Kenney, G., and 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.

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.

Maas, 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.

Macrae, C., Bodenshausen, G., & Milne, A. (1999) Saying no to unwanted thoughts, Self-focus and the regulation of mental life. *Journal of Personality and Social Psychology, 74*, 578-590.

Macrae, C., Bodenshausen, G., Milne, A., & Jetten, J. (1994) Out of mind but back in sight, Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67,* 808-817.

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.

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.

Monin, B., & Miller, D. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81*, 33-43.

Moore, R. (1990). *The formation of a persecuting society, Power and deviance in Western Europe, 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.

Neckerman, J., & Kirschenman, K. (1991). "We'd love to hire them, but…" the meaning of race for employers. In C. Jencks and P. Person (Eds.) *The Urban Underclass* (pp. 203-34). Washington, DC: Brookings Institution.

Nelson, T., Acker, M., & Manis, M. (1996). Irrepressible stereotypes. *Journal of Experimental Social Psychology*, *32,* 13-38.

Nosek, B., Banaji, M., Greenwald, A.(2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics*, *6,* 101-115.

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.

Nunes, A., and 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.

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.

O'Brien, L.,& Crandall, C. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29,* 782-789.

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. (2003). The mark of a criminal record. *American Journal of Sociology, 108*, 937-975.

Peterson, B., Doty, R., & Winter, D. (1993). Authoritarianism and attitudes toward contemporary social issues. *Personality and Social Psychology Bulletin, 19, 174-184.*

Reskin, B., & Bielby, D. (2005). A sociological perspective on gender and career outcomes. *Journal of Economic Perspectives, 19*, 71-86.

Riach, P., and Rich, J. (2002). Field experiments in discrimination in the market place, *The Economic Journal, 112*, F480-F518.

Richeson, J., Trawalter, S., & Shelton, J. (2005). African Americans' implicit racial attitudes and the depletion of executive function after interracial interactions. *Social Cognition, 23, 336-352.*

Roberson, L., & Kulik, C. (2007). Stereotype threat at work. *Academy of Management Perspectives*, *21*, 24-40.

Rodgers III, W. (Ed.) *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.

Shelley, R., & Shelly, A. (2009). Speech content and the emergence of inequality in task groups. *Journal of Social Issues, 65*, 307-333.

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.

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.

Steele, C. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In Berkowitz, L. (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, 9,* 797-811.

Stockdale, M., & Crosby, F. (2004). *The psychology and management of workplace diversity*.

Malden, MA: Blackwell.

Swim, J., & Samma, 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.

Tajfel, H., Billig, M., Bundy, R., & Flament, C. (1971). Social categorisation and intergroup

behaviour. *European Journal of Social Psychology, 1,* 149–178.

Tal, A., Moran, G., Rooth, D., & Bendick, Jr., M. (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*, at www.bls.gov/cps/cpsaat11.pdf, downloaded February 10, 2010.

U.S. Bureau of Labor Statistics (2009). *Job Openings and Labor Turnover-August 2009*, at www.bls.gov/news.release/pdf/jolts.pdf, downloaded October 26, 2009.

U.S. Equal Employment Opportunity Commission (2010). *Charge statistics FY 2007 through FY 2009,* at www.eeoc.gov/eeoc/statistics/enforcement/charges.cfm, downloaded February 22, 2010.

Valian, V. (1998). *Why so slow? The advancement of women.* Cambridge, MA: MIT Press.

Wenneker, C., Wigboldus, D., & Spears, R. (2005). Biased language use in stereotype maintenance: The role of encoding and goals. *Journal of Personality and Social Psychology*, *84*, 470-484.

Word, C., Zamma, M., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interactions. *Journal of Experimental Social Psychology*, *10*, 109-120.

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.

Appendix

A Primer on Matched Pair Tests of Hiring

Matched pair testing (synonyms: situation testing, paired-comparison testing, employment auditing, field experiments, and employment testing) is a systematic research procedure for creating controlled experiments analyzing employers' candid responses to employees' personal characteristics (Bendick, 2007). Economists define employment discrimination as valuation in the labor market of workers' characteristics not related to productivity (Arrow, 1998). In matched pair testing, pairs of research assistants apply for the same actual job vacancy. Within each pair, employee characteristics related to a worker's productivity on the job — such as education, work experience, professional certifications, and technical skills — are made equal 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 only 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 these pairs experience substantially different responses to their job-seeking efforts, few assumptions are required to attribute that difference to the employers' reaction to that characteristic.

Of course, this interpretation 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 situation testing studies that involve only resumes which are mailed, faxed, or e-mailed to employers (Bertrand & Mullainathan, 2002; Bendick, Jackson & Romero, 1996). In these studies, testers' resumes describe equivalent education, work experience, and job skills while varying details and wording to avoid appearing obviously similar. The resumes communicate the

demographic characteristics of the applicants through signals such as gender-specific names, ethnically-related activities, or age-revealing dates of academic degrees, and similar clues. In nations where resumes customarily contain applicant photographs, these images can also communicate demographic characteristics.

Resume-based tests probe only the initial stages of the hiring process, ending with employers' decision to invite job candidates for in-person interviews. To study the complete hiring process, it is necessary to provide "live" testers to file applications, return messages, undergo pre-employment skills tests, and be interviewed. Ensuring that the testers in each pair appear equally 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 simultaneously making accurate observations about the hiring process; willingness to approach the study objectively; similarities between testing partners in general appearance and demeanor; and the differing demographic characteristics required by the study design. Recruiting individuals meeting all the requirements is often a time-consuming, painstaking process; in one typical study, 93 potential testers were interviewed before 4 were selected to form 2 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 to make pairs of testers equally credible job applicants. During training, testers develop and memorize their false resumes, receive coaching on effective job interviewing techniques, and rehearse similar answers to common interview questions. Concurrently, testers are trained to be "human tape recorders" by drilling them to notice and

remember important details of their testing experiences and report them objectively. In typical

well-run testing studies, training requires about 3 days.

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 within minutes of the first. Each tester documents his or her

experiences as soon as practical after the event and prior to being told the experiences of her/his

testing partner. Testers typically record their data using pre-structured questionnaires, and they

are constantly reminded to focus on observable facts rather than interpretations. Such careful

management requires continuous, hands-on monitoring of each test by a trained "Test

Coordinator," who can usually supervise no more than 3 testing teams concurrently.

Testing studies typically repeat their job application quasi-experiment for dozens or

hundreds of job vacancies, to "average out" random circumstances which may affect the

outcome in any single test. In analyzing test 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 (e.g., African American) is successful minus the proportion of

applications in which testers with the characteristic hypothesized to be favored (e.g., Whites) is

successful. Another important summary measure is the ratio of the proportion of tests in which

testers with the disfavored characteristic are successful divided by the proportion of tests in

which the other tester is successful. In both computations, "successful" is typically defined as

reaching an identifiable milestone in the hiring process, such as being offered a job interview or

being offered a job.

Given the 20% to 40% net rates of discrimination observed in typical testing studies,

statistically significant estimates of the main effect studied in the experiment -- e.g., whether the

net rate of discrimination is above zero – can be obtained with as few as 40 to 50 completed

tests. Samples of about 100 tests have sometimes proved sufficient to observe statistically

significant effects of mediating factors on net rates of discrimination – e.g., whether the net rate

of discrimination is higher in one location than another. Analyses that involve multivariate

techniques, such as multiple regression analysis (e.g., Kenney & Wissoker, 1994), often require

larger samples.

       The following are examples of hiring outcomes that testing studies have concluded

indicate bias:

- A large-circulation 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 similar-appearing 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, interviewed the day after that, and

  subsequently offered the position. The African American tester made 4 follow-up calls to

  reiterate his interest, including one shortly after the White tester declined the job offer,

  with no response (Bendick, Jackson & Reinoso, 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

was given an appointment for an interview the following morning (Bendick, Jackson, Reinoso & Hodges, 1991).

- An employment agency advertised for an "account representative" to do executive recruiting. Two white males, whose resumes and appearance portrayed them as age 32 and 57 respectively, responded by telephone and were granted interviews. The older tester's interview lasted 48 minutes, during which the tester 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, during which the interviewer discussed a variety of work and non-work topics in a friendly manner and commented enthusiastically on the tester's questions and 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 who interviewed her told her that "the auto lube job is hard for a woman," said that he liked her smile, and offered an alternative, lower-paying customer service position in the on-premise coffee kiosk. When her male testing partner applied for the position several hours later, he was interviewed for the advertised technician position (Nunes and Seligman, 2000).

The most frequent criticism of testing studies is voiced 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 testing researchers to match the testers on some subtle

but important productivity-related characteristics (e.g., Heckman, 1998; see also Pager, 2007).

Such criticism is tautological, as well as contradicted by the research reviewed throughout this

paper. However, it provides an important caution that only researchers committed to conducting

their research with sustained thought, care, and rigor should undertake matched pair testing.

About the Authors

MARC BENDICK, JR., is a principal in Bendick and Egan Economic Consultants, Inc., in Washington, DC. An economist (Ph.D., 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, a frequent expert witness in employment discrimination litigation, and has led or participated in several dozen matched pair testing studies.

ANA P. NUNES is a Ph.D. candidate in social psychology at the University of Colorado at 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 ran matched pair testing projects investigating housing discrimination for Project Sentinel in Palo Alto, California, and employment discrimination for the Discrimination Research Center in Berkeley, CA.