Recruitment and Bias

Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty
Frances Trix and Carolyn Psenka, *Discourse & Society,* 14 (2): 191-220 (2003)

The Impact of Gender on the Review of the Curricula Vitae of Job Applicants: A National Empirical Study
Rhea E Steinpreis, Katie A Anders, and Dawn Ritzke, *Sex Roles,* 41 (7/8):509-528 (1999)

Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination
Marianne Bertrand and Sendhil Mullainathan, *The American Economic Review*, 94: 991-1013 (2004)

What We Know Now About Bias and Intergroup Conflict, the Problem of the Century
Susan T. Fiske, American Psychological Society

A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance
Claude M. Steele, *American Psychologist,* 52 (6):613-629 (1997)

Reducing Prejudice: Combating Intergroup Biases
John F. Dovidio1 and Samuel L. Gaertner, *Current Directions in Psychological Science*

Avoiding Bias in Reading and Writing Evaluations

The Dual Career Problem
Laurie McNeil and Marc Sher, *Physics Today*, 32-37(1999)

ADVANCE Handbook for Faculty Searches and Hiring
University of Michigan ADVANCE

Are Women Opting Out? Debunking the Myth
Heather Boushey, Center for Economic Policy and Research

Elite male faculty in the life sciences employ fewer women
Jason M. Sheltzer and Joan C. Smith, Proceedings of the National Academy of Sciences

Exploring the color of glass: letters of recommendation for female and male medical faculty

Discourse & Society
Copyright © 2003
SAGE Publications
(London, Thousand Oaks,
CA and New Delhi)
www.sagepublications.com
Vol 14(2): 191–220
[0957-9265
(200303) 14:2;
191–220; 026277]

FRANCES TRIX AND CAROLYN PSENKA
WAYNE STATE UNIVERSITY

ABSTRACT. This study examines over 300 letters of recommendation for medical faculty at a large American medical school in the mid-1990s, using methods from corpus and discourse analysis, with the theoretical perspective of gender schema from cognitive psychology. Letters written for female applicants were found to differ systematically from those written for male applicants in the extremes of length, in the percentages lacking in basic features, in the percentages with doubt raisers (an extended category of negative language, often associated with *apparent commendation*), and in frequency of mention of status terms. Further, the most common semantically grouped possessive phrases referring to female and male applicants ('her teaching,' 'his research') reinforce gender schema that tend to portray women as teachers and students, and men as researchers and professionals.

KEY WORDS: *academic medicine, apparent commendation, discourse analysis, gender bias, letters of recommendation, methodology, possessives*

Gatekeeping practices, including educational requirements, job interviews, and letters of recommendation, all serve to control access to particular positions and the societal benefits that thereby accrue. At the same time gatekeeping practices are all potentially revealing, particularly in times of social change, as institutions replicate themselves and seek to control change. However, studying these practices is challenging, for the higher the social status of the institution, the less public the gatekeeping. We are reminded of Marguerite Yourcenar, the first woman elected to the French Academy since its founding in 1634, and her famous address in 1981 in which she noted that 'the Academy hadn't been particularly misogynist. Rather it had merely conformed to the practice of readily putting women on a pedestal, but not yet allowing them to be offered a seat' (Yourcenar, 1981: 2). Thus, we remember her words, but the gatekeeping

practice through which she was elected is shrouded in the privacy of the 39 male members of the time.

Not just access, but even interest in gatekeeping practices is limited by the way inequities in the advancement of women and minority groups are often hidden by media attention to the few exceptional women or minorities who do succeed in reaching positions of power. These few seem to repudiate allegations of inequity. But as several law cases have argued, people shouldn't have to prove exceptional, only equal to others who were promoted at the same time (Selvin, 1993). Statistics of hiring and promotions in professional institutions, along with studies like the one of the Swedish Medical Research Council's sexism and nepotism in awarding postdoctoral fellowships (Wenneras and Wold, 1997),[1] also call into question gatekeeping practices through which similar people are selected, year in and year out.

In this article we analyze a naturalistic set of all the letters of recommendation for successful applicants for faculty positions in a large American medical school for a three-year period in the mid-1990s. We were asked by a member of the Executive Committee for Hiring and Promotion of the medical school to see if the letters of recommendation written for female applicants were systematically different to those written for male applicants. The broader social context is that of professions in America in which women's greater access to educational opportunities in medicine, law, business, seminaries, and academia since the 1970s has not resulted in a commensurate movement of women into positions of power in these institutions and their related organizations (Valian, 1998).[2]

Specifically in academic medicine in the USA, the institutional context of this study, in the early 1990s women made up close to 20 percent of medical faculty, and their chances of receiving tenure were half that of male colleagues (Brownlee and Pezzullo, 1992). At that time, there were no female deans of medical schools and of 2000 departments, only 85 had female chairs (Brownlee and Pezzullo, 1992). By the mid-1990s, women accounted for 32 percent of the assistant professors, 21 percent of the associate professors, and 10 percent of the full professors (Association of American Medical Colleges, 1996), or 22 percent of medical faculty. What makes this more problematic is the greatly expanded pool for female medical academics. The percentage of females in medical school classes had risen from 8 percent in 1964 to 42 percent in 1994 (Association of American Medical Colleges, 1996: Table B1). And academic medicine as opposed to private medical practice was attractive to a significant number of female medical school graduates. Yet once in academic medicine, women were still taking significantly longer to advance than men.

Even those women who were able to rise in academic medicine reportedly worked under sex-related stress. The resignation in 1991 of Dr Frances Conley from Stanford Medical School, the only female neurosurgeon in her department, brought to public notice what she referred to as 'demeaning actions and words for twenty-five years' (Conley, 1998: 120). In the public discussion that ensued, 'medical professionals of both sexes agreed that academic medicine was a

particular hothouse of sexist attitudes because of the rigid educational hierarchy, the traditional inequality between doctors and nurses which sets the tone for other working relationships, and the many opportunities to make rude anatomical remarks' (Gross, 1991: 10). Female medical academics noted the daily slights that were so wearing. A study at Johns Hopkins University School of Medicine in 1990, found that only 38 percent of its female faculty felt welcomed members of the institution, in contrast to 74 percent of its male faculty; and 75 percent of female faculty felt that men had difficulty taking careers of women faculty seriously and accepting them as colleagues (Fried et al., 1996: Table 2).

American medical colleges are classic examples of male-dominated institutions. Because of the social prestige of medicine in the USA – in the context of academia, medical faculty as a group have the highest salaries – medical colleges have been able to remain more socially conservative in a time of social change. What this implies for letters of recommendation is that gender schema,[3] that is, sets of largely non-conscious assumptions about sex differences in men and women (Valian, 1998), that so affect expectations and interpretations of interactions, may be more overt here. Amplifying this is the convention that letters of recommendation for medical faculty should be written by heads of department. Heads of departments in medical colleges are overwhelmingly older males whose own experience of medical training took place at times when women were largely nurses, patients, or stay-at-home wives.

In contrast to the data of much discourse analysis of racism or prejudice, where the Other is already a group whose agency is backgrounded or suppressed (van Leeuwen, 1996) and whose individuality has been reduced or is nonexistent, here we begin with letters for individuals. Theoretically then, the letters for women, or foreigners or African Americans had we sufficient numbers, potentially show the process of reduction from individual to Other. In terms of our framework, the least persuasive letters for female applicants describe them in ways that ignore or downplay their professional accomplishments and individual qualities, reducing them to gender schema that see women as less capable and less professional in the demanding work of academic medicine.

From a meta-research point of view, we are thus examining a situation in which contrasts will most likely be greater and more obvious than in letters of recommendation for other areas of academia. Like Goffman (1981) in his study of radio talk, in which he built upon its restricted interactional setting – that of the radio announcer who lacks direct feedback – to then propose ways of looking at more complex face-to-face interaction, we too hope to propose ways of thinking and methods of research that could be applied to data with more subtle distinctions. Although the gender schema that affect both men and women throughout society tend to be unarticulated, examination of these letters of recommendation should provide another way of studying such schema where there is explicit language and meaningful contrasts, along with the better known statistics, hearings (Trix and Sankar, 1998), and law cases. Methodologically this study also wrestles with a situation in which what is not written is potentially

salient. Thus the 'glass' in the title refers both to the 'glass ceiling' that appears to be impeding women from advancing professionally, as well as 'glass' or invisible domains in the letters themselves.

Previous research on letters of recommendation

Studies of letters of recommendation have come largely from the fields of education (Morisett, 1935; Rayber, 1985), psychology (Cowan and Kasen, 1984; Hatcher, 1983), English (Eger, 1991), sociology (Bell et al., 1992), linguistics (Bouton, 1995; Precht, 1998; Watson, 1987), business (Tommasi et al., 1998), and medicine (DeLisa et al., 1994; Gayed, 1991; Greenburg et al., 1994; Johnson et al., 1998; O'Halloran et al., 1993). Besides these systematic studies, there are also numerous short articles in multiple fields that reflect anecdotally or give advice on writing letters of recommendation. In the systematic studies, the main concerns are reliability (Hatcher, 1983; Morisett, 1935; Tommasi et al., 1998), relative importance of letters of recommendation for screening candidates (DeLisa et al., 1994), deficiencies in letters of recommendation (O'Halloran et al., 1993), cross-cultural differences and evaluations (Bouton, 1995; Gayed, 1991; Precht, 1998), identification of features linked to positive and negative interpretations (Greenburg et al., 1994), and sex-linked features or patterns (Bell et al., 1992; Eger, 1991; Hatcher, 1983; Watson, 1987). The features most commonly studied are: length, naming practices and gender identification, negative language, and sex-linked descriptive terms.

An unusual study of particular relevance is that of Eger (1991) who took a psychological approach and profiled 12 male and 12 female writers of letters of recommendation according to the Myers Briggs personality assessment. He then gave each of the writers files on six applicants that contrasted in gender and personality for whom to write recommendations. Eger hypothesized that people would write better letters for people who were more like themselves in personality, ideology, and gender. This he indeed found to be true. This 'advocacy factor' as he called it, has relevance for this study in that letters of recommendation for medical faculty positions are overwhelmingly written by men.

Among studies of sex-linked features, Watson (1987) analyzed 80 letters of recommendation for graduate study in social sciences (40 letters written by males: half for males, half for females; 40 letters written by females: half for males, half for females). She found that the longest letters were written by female recommenders for female applicants, that men tended to use gender identifications with female applicants twice as often as females did for male applicants, that females used first names of females more frequently than males did, and that comments about appearance and condescending adjectives were used only for female applicants, whereas only men were praised for their sense of humor. We relate some of these to the findings for letters of recommendation for medical faculty. But in our study the contrast of female and male recommenders was not possible because such a relatively small proportion of the recommenders were

female. We wonder if the recommenders in Watson's study were of equal status as well.

Another study of sex-linked features, not systematic but nevertheless stimulating, is LaCroix's on gender bias in recommendations for admission to colleges (LaCroix, 1985). LaCroix noted frequent stringing together of stereotypical terms, like 'Betty is a polite, quiet, gentle, friendly, and cooperative member of her class' (LaCroix, 1985: 25). What is most important is not that these terms are negative in themselves, but that they take the place of more substantive comments about academic characteristics. Thus, there were 'knowledge gaps'. This we also found in letters of recommendation for medical faculty. But further, one of the most challenging features of letters of recommendation for medical faculty is the growing tendency not to state the negative, but merely to fail to state the positive. Perhaps fear of lawsuits encourages this. So readers need to be carefully attuned to what is not said. This is less problematic when the data are not numerous. But when there are over 300 letters, as in our study, seeking the unwritten requires other methods besides close reading.

As for the studies in medicine, they are primarily focused on the letters of recommendation used to select and place students in residency training programs. We found no previous studies on letters of recommendation for medical faculty. Nevertheless, of the many studies of letters for medical residency programs, the most relevant is that of Greenburg et al. (1994) for surgical residencies. It was based on 80 letters drawn from an applicant pool of over 300 applications for 6 places per year. The letters were evaluated into groups of acceptable, mediocre, and poor, and then different combinations sent out to surgeons across the country for their evaluations. Half of the letters were sent with their original letterheads, half without. Despite surgeons thinking that school of origin mattered, it did not turn out to be statistically significant. As for length, the top five letters were twice as long as the bottom five letters. The top letters also had three times the number of personal references as did the poorly ranked letters. The worst ranked letters either contained no information or ambiguous information. They lacked specificity. The researchers also found that two of the letters ranked as lowest in their group were written on behalf of students whose other letters were in the highest group, thereby confirming the need for multiple letters. Finally, the researchers noted that although the rankings largely coincided, there may be 'code phrases' or understandings that as of yet have not been specified but which affected rankings, especially between the acceptable and the mediocre-ranked letters. Interestingly, there was no mention of gender in the study.

Data and methodology

As noted, the data for this study are a naturalistic set of all the letters of recommendation for successful applicants for faculty positions at a large American medical school over a three-year period from 1992 to 1995. This amounted to 312 letters for applicants for 103 faculty positions, with sets of approximately

three letters per applicant. For legal reasons, only the letters of successful applicants were available for study. The positions applied for were both clinical and research, largely at the assistant professor level, but also including adjunct positions as well as eight associate professorships and one full professorship. Thus, another way to think about the data set is to see it as the potential pool for promotion, including those who will and will not be promoted. The medical specialties were a broad spectrum of 37 different ones, including surgery, oncology, neurology, internal medicine, obstetrics–gynecology, urology, radiology, pathology, psychiatry, pediatrics, family medicine, and anatomy. We attempted to group these specialties but in the end gave up. Apart from gross differences of pay and prestige between surgery, internal medicine, and basic science, physicians could not agree on the categories.

As for the gender of the applicants, 89 of the letters were for women, 222 of the letters were for men, and one letter was for a couple. That is, 29 percent of the letters were for women, whereas 71 percent of the letters were for men. (Unfortunately we cannot know the gender breakdown of total applicants, including the unsuccessful ones.) The overall percentages of men and women at the different ranks at this large medical school were similar to the national average, so we assume this gender contrast in hiring is not unusual. Specifically, at this medical school in 1995, 31 percent of the assistant professors were female, 19 percent of the associate professors were female, and 10 percent of the full professors were female. Overall, 23 percent of the medical faculty were female.

As for the recommenders, they were 85 percent male, 12 percent female, and 3 percent unknown (probably foreign names that could not be readily categorized by the office that blacked out the names). The high proportion of male recommenders reflects the convention of having heads of department, who are overwhelmingly male, write the letters. The institutions from which the recommenders hailed were university teaching hospitals and major urban hospitals across the USA, Canada, and overseas. We attempted to categorize these institutions out of concern for potential differences of familiarity and culture. We grouped the places of origin into four categories: United States – local (215 letters), United States – not local (79 letters), Canada, Britain, Australia, South Africa, Israel or CEASI (12 letters), and Europe but not Britain (5 letters). By 'local' we refer to the geographic region surrounding the medical school in question, comprised of areas of three adjoining states. We assumed that physicians in this proximity in the same specialties would know each other and their letters might differ from those who didn't know each other. The CEASI group is a largely English and American-oriented group. The problem with this classification is that recommenders from different cultural backgrounds could be in any of these locations, and we could not know how long they had been in their current institutions.

Nevertheless, we did notice that letters from Europe tended to be shorter. Even letters from Canada were less hyperbolic than those from the USA. But we did not have enough letters to make more than general observations. Anecdotally, we

have been told that in Germany, for example, the stature of the recommender matters more than carefully constructed content. This may also have been true earlier in America when physicians knew all the main people in their fields and there were many fewer hospitals and medical schools. We were told by older physicians that in the 1940s, phone calls alone were often sufficient for providing recommendations. This suggests a more unitary and informal structure of decision-making, one fraught with potential misuse by today's standards.

The gatekeepers, that is, the people to whom the letters were addressed, were overwhelmingly male. Specifically, 96 percent of the gatekeepers were male, 4 percent female. Combined with the high percentage of male recommenders (85%), the letters can be seen as largely in-group discourse, written by male physicians to male physicians. In keeping with recent research on institutional discourse, this is discourse between professionals (Drew and Sorjonen, 1997); in social terms it is discourse between elites (van Dijk, 1993). The forms of address of the gatekeepers are also interesting. Most gatekeepers were addressed formally as 'Dear Dr Koop' (whatever last name). However, 40 were addressed, 'Dear Al' (whatever first name). The gatekeepers addressed by first name were all males, constituting 13 percent of the letters to male gatekeepers, and reinforcing the in-group nature of the letters.

After we had conducted the initial analysis of the letters, we took sets of nine letters that we had rated high, average, and deficient, and showed them to physicians of different specialties across the region. We asked the physicians to rank the letters and comment on them. Their rankings were in line with ours; a few also commented on subtexts that we had not noticed. For example, one physician noted after reading the closing paragraph of a letter we had all ranked high, that the recommender had also implied the candidate was good enough for the proposed institution but not for his own. Interviews with physicians reinforced our understanding of the strong hierarchical consciousness of the American medical profession.

We then analyzed the letters according to the traditional categories of length, naming practices, negative language, and sex-linked terms. Through interacting with the data we modified and expanded the category of negative language, tabulated differences in repetition of status terms across the letters, and added a category of letters lacking in basic features. Our most interesting new category is that of semantic realms following possessives.

We want to emphasize that the exigencies of the original audience for our research, that is, medical academics who are wedded to statistics as an essential tool of persuasive research, in combination with our adherence to naturalistic data, and the gender disparity of successful applicants in the field, led us to request large numbers of letters so that the percentages of letters for women could be statistically significant. The resulting corpus was so large that it initially required analysis at a broader level than many discourse studies. We feel that the combination of critical discourse analysis with methods for corpus analysis will increase the potential audience and effect of discourse research.

General findings

One of the simplest variables in ranking letters of recommendation is *length*. As the structure of these letters is fairly conventional with an introductory section wherein the recommender notes the relationship with the applicant, a body where academic traits and achievements are noted, and a closing section where the recommendation is made (Bouton, 1995: 221), a letter that is excessively brief may omit some of these expected elements. A secondary pattern, particularly in letters for promotion or higher rank entry, involves an expansion of the body to include evidence of productivity in research, effectiveness in teaching, and collegiality in service. Often the more detail in the letter, the more persuasive. A longer letter shows care on the part of the recommender at the very least. Whereas a very brief letter may be the result of laziness on the part of the recommender – the result of not knowing the applicant well and not taking the time to read the applicant's résumé – or it may be that there is nothing positive to say. Whatever the level of responsibility or writing skills of the recommender, the effects are interpreted as reflecting on the applicant.

By number of words, the average length of the letters was 246 words, with the average length of letters for female applicants being 227 words, whereas the average length of letters for male applicants was 253 words. To help visualize length, we standardized all the letters to have lines of 97 characters including spaces. Thus we were able to compare letters in terms of lines as well as numbers of words. As a convention we also included the salutation and closing as separate whole lines. We found great similarity in length of letters for female and male applicants in the mid-ranges of length of 11–20 lines or 111–202 words (45% of letters for women, 44% of letters for men), 21–30 lines or 212–303 words (25% of letters for women, 25% of letters for men), and 31–50 lines or 313–505 words (19% of letters for women, 15% of letters for men). However, at the extremes there were differences. At the high end, 8 percent of the letters for men were over 50 lines long, whereas only 2 percent of the letters for women were this long. The longest letter of our data was written by a female recommender, but unlike Watson's results, for a male applicant (116 lines, 1354 words). Whereas at the low end, 10 percent of the letters for women were 10 lines (111 words) or fewer, whereas 6 percent of the letters for men were this short. Short letters reflect to a certain extent the inclusion of letters for adjunct faculty, but they do not explain the gender discrepancy for there were over three times the number of successful male applicants for adjunct positions as successful females, and some letters for assistant professorships for women were equally scant.

In studying the shorter letters further, we found a whole new class of letters that we prefer to term *letters of minimal assurance* rather than letters of recommendation, as they were lacking in relevant features. The following is an example of such a letter:

TEXT 1. EXAMPLE OF *LETTER OF MINIMAL ASSURANCE*

Dear Dr Alfred Koop:

It gives me great pleasure in writing this recommendation letter for Dr Sarah Gray. I have known Sarah as a resident and as staff at Mrahonod Metropolitan Hospital. She is knowledgeable, pleasant, and easy to get along with. I have no hesitation in recommending her for a faculty position at Centvingcinq.[4] I will be happy to answer any further questions in this regard.

Charles Lewis, MD
Chairman, Department of Psychiatry

From our analysis of the letters, and supported by the literature (Johnson et al., 1998: 42), we expect letters of recommendation to include: commitment and relationship of recommender with the applicant, some specificity of focus and record of the applicant, and some evaluation or comparison of traits and accomplishments of the applicant. Notice that the above letter (Text 1) shows commitment and relationship of recommender with applicant, but no specificity of focus or record, and no evaluation of traits or accomplishments. Where at least one of these three areas was not represented, we considered the letters to be of *letters of minimal assurance* rather than letters of recommendation.

TEXT 2. EXAMPLE OF *LETTER OF MINIMAL ASSURANCE*

RE: Sarah Gray, MD

I am writing in support of Sarah Gray MD's application for the position of Associate Professor of Nephrology in your department. I have worked closely with Dr Gray both as her chairman and as a fellow faculty member doing pediatric nephrology for the past three years. She is a superb clinician and academician. I truly enjoyed working with her. Your gain is my loss. I believe that you will find that she will be a genuine adjunct to your faculty. If you require more specific information, please do not hesitate to notify me.

Sincerely,
Charles Lewis, MD
Professor & Chairman, Depart. of Nephrology

In the above letter (Text 2), the commitment and evaluative comments are present. What is problematic is specificity of focus and record of the applicant. Notice that it is the recommender who is the agent in 'doing pediatric nephrology', although the applicant has a backgrounded role in this as 'fellow faculty member.' The lack of specificity of the applicant's record is particularly egregious as the position is for Associate Professor. It appears that the recommender senses this, and therefore suggests that more specific information could be forthcoming, were it required.

In the corpus, we found systematic differences in the relative percentages of *letters of minimal assurance*. Fully 15 percent of the letters for female applicants fit into this category of letters of minimal assurance, whereas only 6 percent of the letters for male applicants fit into this mold (Figure 1). This is statistically significant with a *p*-value of .021.

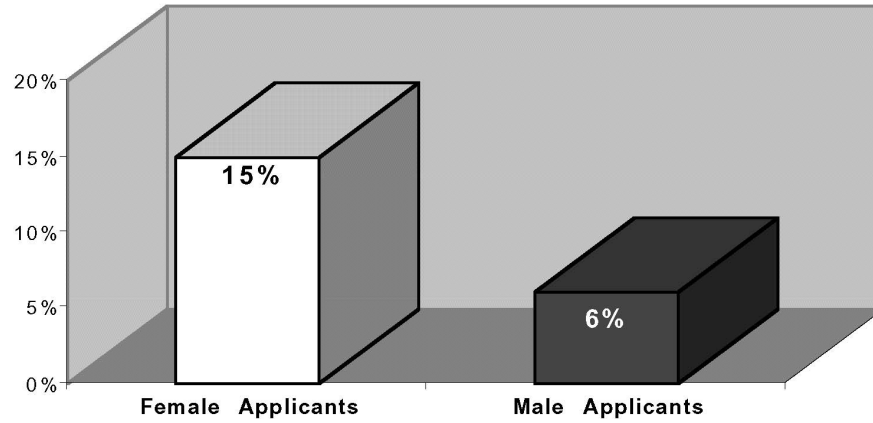


FIGURE 1. *Percentage of letters of minimal assurance*

Although such letters of minimal assurance were more common among the shortest letters, generally 70–100 words, length alone was not the defining feature. (As mentioned, the average length of all letters was 246 words.) For example, the following two letters are very short and yet include the relevant areas as described above.

TEXT 3. EXAMPLE OF SHORT LETTER OF RECOMMENDATION

> RE: Appointment of Sarah Gray, MD:
>
> I am pleased to recommend Dr Sarah Gray for faculty appointment as Clinical Assistant Professor. I have known Dr Gray for 8 years. *She worked* in research with me for 1 year and *did fellowship training* in our program for 2 years. She is a very good internist and endocrinologist. She is honest and reliable and of highest moral quality. She has good judgment in patient care and is very thoughtful and considerate towards those she is caring for. She is a good clinical teacher and should serve the department well in the capacity of instructing students and residents.
>
> FROM: Charles Lewis, MD, Director, Endocrine and Hypertension Division

All the features of commitment of recommender, and specificity and evaluative comments regarding the applicant are present in this memo-type letter (Text 3). Note as well the second sentence where the applicant is portrayed as the agent: 'she worked in research and did fellowship training.' Such agency foregrounds the applicant as an active person in her own right.

TEXT 4. EXAMPLE OF SHORT LETTER OF RECOMMENDATION

> Dear Dr Koop:
>
> I am writing to support Dr Harvey's appointment as Clinical Assistant Professor of Medicine. Dr Harvey did his Fellowship here and I got to know him well. I

consider him highly intelligent, highly motivated, and highly productive (*he wrote* an extraordinary number of papers for a Fellow). He has a thorough knowledge of medicine. He is excellent with patients and is exceptionally pleasant without a shred of egotism, and was highly admired, liked, and respected by his colleagues. I recommend him highly.

Sincerely,
Dr Charles Lewis, MD
Director, Cancer Center

Again, despite it being a short letter (Text 4), the features of commitment of recommender, some specificity of record, and evaluative comments of the applicant are all present. As with the previous short letter of recommendation, notice again that within the parentheses of the third sentence, the applicant is the agent: 'he wrote an extraordinary number of papers'. This productivity stands in interesting contrast to the applicant's being 'without a shred of egotism', the latter not a common trait of male academic physicians.

Another area that has been studied in research on letters of recommendation is *naming practices*. In its simplest form, we did not find differences in this area between letters written for male and female applicants. In general, the first sentence of a letter referred to the applicant in full name or in Dr +last name, as did the sentence in the closing with the recommendation. In the body of the letter, the applicant was referred to variously as Dr +last name, or full name, or first name only. But the percentages of all these uses were similar across the data.

Another sort of naming practice which did show difference was the use of gender terms. As would be expected in the use of gender terms (woman, lady, mother, wife; man, gentleman, father), those for females were more marked.

TEXT 5. EXAMPLES OF USE OF GENDER TERMS IN DIFFERENT LETTERS

Dr Gray is a thorough, hardworking, extremely intelligent and insightful *woman*.
She is an extremely intelligent young *lady* with an admirable work ethic.
On a personal level Sarah is, in my opinion, the quintessence of the contemporary *lady* physician who very ably combines dedication, intelligence, idealism, compassion and responsibility without compromise.
I believe Dr Harvey to be a *man* of great personal integrity.
Overall, we have found William a highly intelligent and hard working young *man*.
He is entirely dedicated to patient care, personable, a *gentleman* in every sense of the word.

The above examples of use of gender terms (Text 5) all come from different letters, generally in the last third of the letter. Note the pairing of 'intelligent' or 'intelligence' with the female gender terms, as if this were remarkable, as well as the clichéd phrases with the male gender terms, 'a man of great personal integrity', and 'a gentleman in every sense of the word'. Of the letters for female applicants, 10 percent included such references, all but one by male recommenders, whereas only 5 percent of the letters for male applicants included use of such

gender terms. Other researchers have interpreted such usage to mean that the writer sees the applicant first in terms of gender and secondarily in terms of position (Watson, 1987: 27).

But we would like to add a fuller context for the unusual 'quintessence of the contemporary lady physician' phrase. Most female physicians would not like to be referred to as 'lady physicians', but we believe this reflects a recommender from an older generation. Further, we see the recommender of this letter effectively supporting the appointment of the applicant, with a subtext of indirect criticism of many contemporary physicians who lack the qualities of this 'lady physician'. The letter is the third longest for a woman, 44 lines or 504 words, made up of six paragraphs. In the first five paragraphs, the female applicant is referred to as Dr Gray. She is also described as 'Staff Physician', with her record of post-residency training and work (paragraph 1), 'an individual dedicated to her work' who has mastered multiple difficult procedures (paragraph 2), 'a refreshing type of person' in that she tends to underrate herself but who actually operates at a sophisticated level of problem solving and expertise (paragraph 3), 'the quintessence of the contemporary lady physician' who is held in high esteem by those who work with her at all levels (paragraph 4), and 'a former trainee' to whom the recommender gives unqualified endorsement for advancement (paragraph 5). In the sixth and final paragraph the applicant is referred to by her first name, and surprisingly also as 'my surrogate' (paragraph 6). For a male recommender to see a female applicant as 'his surrogate' crosses gender lines and is truly worthy of note and hope.

The other naming practice that showed distinctiveness was use of titles such as 'Chief Resident' or 'Head of pediatric cardiology'. Only 3 percent of female applicants' letters referred to them with a particular title other than 'Dr', whereas 12 percent of male applicants' letters included such titles. The problem here is knowing if more women had titles and they were not used, as was the case in two of the three letters of all the sets of letters that included use of a woman's title, or if indeed there was a higher percentage of men who had achieved the prominence to be accorded such titles.

A third area that has been discussed, especially in the non-systematic articles on letters of recommendation, is *negative language*. In letters of recommendation, negative comments stand out like sore thumbs. Some recommenders feel they will be more credible if they also include negative points, but the result is usually that the negative comments are more memorable. In coding the 312 letters, we came to theorize that negative language was part of a larger set of elements that potentially raised doubt in the mind of the evaluator. Similarly, another study noted signs of 'hesitancy' on the part of the recommender and faint praise (Greenburg et al., 1994: 197). We defined these *doubt raisers* as including negative language, along with hedges, potentially negative comments, unexplained comments, faint praise, and irrelevancies. In general, *doubt raisers* were found in the middle section of letters.

TEXT 6. EXAMPLES OF CATEGORIES OF *DOUBT RAISERS*

negative language

While Sarah has not done
a lot of bench type research,
She has a somewhat challenging personality.

Although his publications are not
numerous as you know,
While not the best student I have had,

hedges

It appears that her health and personal life
are stable.

He appears to be a highly motivated
colleague,

potentially negative

As an independent worker she requires only
a minimum amount of supervision.

Bright, enthusiastic, he responds well to
feedback.

unexplained

Now that she has chosen to leave the laboratory.

faint praise

She worked hard on projects that she
accepted.

I have every confidence that Bill will
become better than average.
He is void of mood swings and temper
tantrums.

irrelevancy

She is quite close to my wife.

He is very active in church.

Irrelevancies are not innately negative, but the overall effect can raise doubt for it appears that the recommender could find nothing more substantive to say about the applicant.

Again, with respect to the category of *doubt raisers*, there were systematic differences between letters written for male and female applicants. Specifically, whereas 24 percent of the letters written for female applicants had at least one *doubt raiser*, only 12 percent of the letters written for male applicants had at least one *doubt raiser* (Figure 2). This is statistically significant with a *p*-value of .01.

Further, some letters have more than one *doubt raiser*. Of the letters for female applicants with *doubt raisers*, there is an average of 1.7 per letter. Of the letters for male applicants with *doubt raisers*, there is an average of 1.3 per letter. Such multiple *doubt raisers* can build on each other.

For example, the following letter frames the applicant early on with an irrelevancy, namely that the applicant is 'quite close to my wife and they frequently seek each other's company.' The interpretation could develop that the recommender wrote the letter of recommendation because his wife insisted he do so. Whatever, it gives the applicant a decidedly less than professional alignment. This worsens with the second *doubt raiser*, namely the applicant's unexplained choice to leave the laboratory. Laboratory research has higher status than clinical work and so there needs to be either some positive comment on the laboratory work the applicant did previously, or an explanation of why she is focusing on clinical work, or no mention of the change.

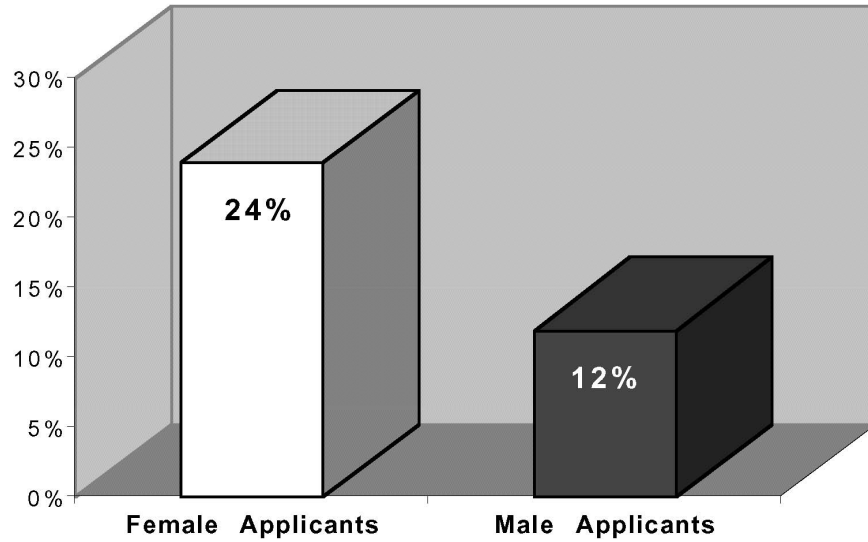


FIGURE 2. *Percentage of letters with* doubt raisers *(negative language, hedges, potentially negative, unexplained, faint praise, irrelevancies)*

TEXT 7. LETTER WITH MULTIPLE *DOUBT RAISERS*

Dear Alfred:

It is a pleasure to write a letter in support of Sarah Gray MD, one of our urologists who is leaving St Louis and would join your program. I have known Sarah for approximately 4 years, becoming socially friendly with her and her husband, particularly over the past year and a half. *Sarah is quite close to my wife, and they frequently seek each other's company out*. Obviously, Sarah is a very committed physician who has very good clinical skills, ambition, and a desire to participate actively in the care of children with urologic problems. I get a sense that she is looking to develop a clinical program *now that she has chosen to leave the laboratory* and concentrate exclusively on clinical medicine.

Although I can't specifically comment on Dr Gray's clinical skills in detail, I have been impressed with her care of patients that we are mutually involved with. I believe she is a concerned and interested clinician who offers excellent care, *tries hard to communicate* with the patients and with the physicians. *She is more academic than most clinical physicians* and this should be a resource as she becomes involved in her own programs.

Although *we will miss Sarah* we are sure she will be a great asset to your program and, therefore, *we wish her well*.

If there are further questions which I can answer regarding Dr Gray please do not hesitate to call.

Yours sincerely,
Charles Lewis, MD
Professor of Pediatrics,
Chief, Division of Urology

Notice as well the hedges on the part of the recommender, such as, the applicant 'tries hard to communicate with the patients and the physicians.' Is she not also a physician? And with the fourth italicized comment, 'she is more academic than most clinical physicians,' one wonders in what respect? In the next statement, 'we will miss Sarah,' which 'we' is being referred to – the hospital staff or the recommender and his wife? Overall, the commitment on the part of the recommender is not convincing. Why does the recommender not know for sure why she has left the laboratory? And notice that the applicant has no agency other than having chosen to leave the laboratory and concentrate on clinical medicine. The early framing in social rather than professional terms, added to the following hedges and wishy-washy phrasing, make this more a polite acknowledgement of departure – 'we wish her well' – than a recommendation of ability and professional accomplishment.

Another sort of letter with multiple *doubt raisers* has them escalating toward the end of the letter. The longest letter for a woman, nine paragraphs of 80 lines, totaling 849 words, has such an escalation in the last three paragraphs of the letter. Before that, the structure of the letter is a common one of evaluation of teaching, research, administrative record, national contribution, and service. There are occasional *doubt raisers* in these, but they are well blanketed in solid positive comments of ability and performance. In teaching, the applicant 'received an outstanding record with medical and pulmonary medicine students.' In research, the applicant 'worked hard on projects that she accepted, and showed a good ability both as a primary investigator and, more often, as a collaborator' – two examples of faint praise in that the first implies she may not have done all projects expected of her, and the second downgrades her, as being a primary investigator is expected of an associate professor. As an administrator, the applicant 'performed very well', 'served as chair of the appointments committee', and 'established a pediatric pulmonary service'. The one *doubt raiser* here was that 'She has a somewhat challenging personality, but I felt she was appropriate in her manner of confronting issues.' For a woman to have such a personality stands out for it goes against gender schema; even for a man it would not be good. Perhaps the recommender could have written that 'as an administrator she confronts issues directly', and only positive associations would have accrued. At the national level, 'she served as chair of several committees', 'was active in a national study group', and 'is well known and respected'. In the local hospital, she took her fair share of rotations. So far this is still a strong letter.

But the last three paragraphs change the image. I include the seventh paragraph in entirety, with specifics altered to maintain anonymity.

TEXT 8. PARAGRAPH OF INAPPROPRIATE *DOUBT RAISERS*

> [six paragraphs precede]
>
> Sarah's *personal life was in turmoil* during the time I worked with her, and in view of the difficulties she was experiencing in that arena, her performance was

> especially impressive. Her last years in my laboratory were impacted by *serious health problems* that have fortunately gone away – she had really debilitating problems with a herniated disk that apparently was a paraneoplastic phenomenon that went away once an early carcinoma of the left ovary was identified and removed.
>
> [two paragraphs follow]

Similar to van Dijk's *apparent denial* and *apparent concession* moves in racist discourse (van Dijk et al., 1997: 170), this is a move of *apparent commendation*.[5] That is, 'her performance was especially impressive, considering problems of which I have special knowledge'. The recommender then lists the problems in detail and to the detriment of the applicant. This is also wholly inappropriate and even unethical, for the problems are private information. Further, in the recommender's description of the applicant's health problems, she has become a patient, not a potential associate professor of pulmonary medicine. This recalls 19th century gender schema in which women were considered too emotionally and physically fragile to possibly become physicians.

The final paragraph (Text 9) is not redeeming. It brings *apparent commendation* to a high art. Here the recommender reveals that previously the applicant did not receive promotion to associate professor at his university, Northsouthern University, although he supported her then too. This might be inferred from her curriculum vitae, but to state it in the final paragraph is to frame the final recommendation with negative associations. The second sentence is also gratuitous – 'While she has not been able to accomplish a lot in academic pulmonary medicine during the past few years due to career changes and other personal issues, she has continued to grow and mature.' This commendation positions her in terms usually reserved for children. And the final blow, 'her great gift for teaching' is diminished by the qualifier, that it is 'especially in small groups and one on one'. This sort of teaching is often seen as less demanding than the large lecture and grand rounds, more suited to women's 'more private sphere'.

TEXT 9. FINAL PARAGRAPH OF *DOUBT RAISERS*

> [eight paragraphs precede]
>
> I strongly support*ed* Sarah's application for an *Associate Professor appointment at Northsouthern University*, and feel that at the present time she is even more qualified for that level. *While she has not been able to accomplish a lot in academic pulmonary medicine during the past few years due to career changes and other personal issues*, she has continued to grow and mature, and am sure she would be an even stronger and more effective member of your department than she was in St. Louis. Her great gift for teaching, *especially in small groups and one on one*, is something that I know you will come to appreciate. I recommend her without reservation for this proposed appointment.

Even the final sentence in the recommendation, 'I recommend her without reservation for this proposed appointment,' rings hollow, for the last third of the letter, including this final paragraph, would readily foster multiple reservations.

We are heartened that the applicant was hired despite this, but we do not know at what rank. And note, this letter will remain in the applicant's file throughout her stay at the university.

Earlier studies had also looked at *stereotypical terms* associated with particular groups. We came up with a list of descriptors and nouns that were differentially associated with male or female applicants. For the men, the adjective 'successful' (in 7% of letters for men; in 3% of letters for women) and the nouns 'accomplishment' and 'achievement' (in 13% of the letters for men; in 3% of the letters for women) stood out. Not surprisingly, for female applicants 'compassionate' and 'relates well to patients and staff at all levels' stood out (in 16% of the letters for women, in 4% of the letters for men). Notice how well these terms accord with gender schemas that see men as successful and women as nurturing.

TEXT 10. EXAMPLES OF STEREOTYPIC ADJECTIVES IN DIFFERENT LETTERS

> Dr (William) Harvey has been very *successful* in obtaining grants from both the NIH (National Institute of Health) and industry and has developed an excellent clinical trials group for the study and treatment of AIDS.
>
> His substantial record of publication in reputable journals, during those developing years, certainly attests to his overall research capabilities and *accomplishments*.
>
> Dr (Sarah) Gray is a caring, *compassionate* physician who has excellent interpersonal relationships with patients and their families as well as nursing and medical staff.

In further studying the letters for female applicants, we developed a list of adjectives that we term '*grindstone adjectives*', as in putting one's shoulder to the grindstone. These include: 'hardworking', 'conscientious', 'dependable', 'meticulous', 'thorough', 'diligent', 'dedicated', and 'careful'.

TEXT 11. GRINDSTONE ADJECTIVES IN DIFFERENT LETTERS

> She is an extremely *conscientious and meticulous* researcher who devotes her time to laboratory work and the training of graduate students in laboratory technique.
>
> She is a superb experimentalist – very well organized, *thorough and careful* in her approach to research.
>
> I have found William to be *hard-working, thorough, and conscientious* in providing all aspects of patient care.

Again, there is nothing wrong with these qualities. We include the second example of the 'superb experimentalist', from the strongest letter for a female applicant, to show that grindstone adjectives, when well contextualized, can be most impressive. This is not their usual role however. Of the letters for female applicants, 34 percent included grindstone adjectives, whereas 23 percent of the letters for male applicants included them. There is an insidious gender schema that associates effort with women, and ability with men in professional areas.

According to this schema, women are hard-working because they must compensate for lack of ability (Valian, 1998: 170).

Findings related to repetition

Continuing the analysis of descriptors, we developed a list of '*standout adjectives*', namely 'excellent', 'superb', 'outstanding', 'unique', 'exceptional', and 'unparalleled'. In tabulating the percentages of letters for female applicants and male applicants that included these terms, we found them to be similar (63% for women and 58% for men). And yet the letters for men read differently. Our gut reaction after reading many of these letters was that the men had been praised more highly with these terms than had the women. Thus we were led to consider frequency. That is, instead of coding mere occurrence of at least one of these terms in a letter, we coded for multiple occurrences. Here we found that the letters for women that had at least one of these terms had an average of 1.5 terms, whereas the letters for men that included at least one had an average of 2.0 such terms. That is, there was repetition of standout adjectives within men's letters to a greater extent. Below are examples of the greatest numbers of standout adjectives in letters for female and male applicants of similar level and focus.

TEXT 12. GROUPINGS OF STANDOUT ADJECTIVES IN LETTERS

letter for female: (entry level)	Sarah was an *outstanding* resident she is an *outstanding* young woman she is an *excellent* psychiatrist
letter for female: (mid level clinical)	her *superb* performance as a teacher she did a *superb* job of establishing an outreach component she had *excellent* relations with clinicians
letter for male: (entry level)	William has an *excellent* publication record, which includes publications in *excellent* journals including *Science* William is an *excellent* writer he has an *excellent* background and understanding he will make an *excellent* teacher he will prove himself an *excellent* researcher and teacher
letter for male: (mid level clinical)	Dr Harvey is an *outstanding* clinician and teacher his background includes *outstanding* training experience his dedication to patient care and *outstanding* teaching attributes one of the *finest* clinicians I have known Dr Harvey is obviously a *superb* teacher Dr Harvey is an *outstanding* internist and cardiologist

Notice the repetition of the same standout adjectives in the above groupings as well.

Repetition in language is an intriguing feature. In written language it has been studied most in the context of poetry, and can be seen as a simple form of parallelism. In the last years of his life Jacobson was fascinated with parallelism in all

forms of language. Linguists who work with oral language have long accorded repetition respect, from Hymes' work on repetitions and contrasts as a definition of structure, to Bollinger's on prepatterned expressions. (See Tannen, 1989: 36–97, for a thorough summary of discourse work on repetition.) More recently, Ferrara (1994) worked on repetition in psychotherapeutic discourse, including echoing, mirroring, and repeating. All studies note the function of repetition as a cohesive device, and rhetoricians note its persuasive value. However, we had not originally considered repetition as a feature for analysis for letters of recommendation, under the assumption that repetition would tend to 'bleed' meaning, that is, multiple uses would drain particular words of their power. This did not turn out to be the case.

For example, in any academic environment the term 'research' is highly charged. Advancement often hinges on getting research published in appropriate journals, or today on securing large research grants that pay the university generous overhead amounts. Again we found that roughly the same personage of letters for women (48%) as for men (50%) mentioned 'research' at least once. Tabulating *multiple mentions of 'research'* within letters produced another picture however.

Within letters, 35 percent of those for women that mentioned 'research' at least once, mentioned it multiple times. In contrast, 62 percent of the letters for men that mentioned 'research' at least once, mentioned it multiple times. The range of numbers of repetitions varied too, from 2 repetitions to 7 at most in letters for women, whereas among letters for men, many had 4 repetitions, with others ranging as high as 11 repetitions.

There was also an interesting difference in the way in which 'research' was portrayed in the letters. Some focused on the particular research of the applicant, some on general abilities of the applicant in research design, whereas others discoursed on the contribution of the applicant to the research environment of the laboratory or department. Ideally, an applicant would have letters that presented all these perspectives. The following excerpts exemplify general research abilities and contribution to the research environment. They are drawn from a letter for a female researcher who had among the strongest letters for females in the corpus. Notice as well the point about establishing her own research program – an essential ability for a serious researcher.

TEXT 13. STATUS TERM 'RESEARCH' IN TWO PERSPECTIVES

> [one paragraph precedes]
>
> Sarah is an extremely diligent and productive scientist. She is a superb experimentalist – very well organized, thorough and careful in her approach to *research*. She knows what questions to ask and how to design sound experiments to answer them. She is perhaps unusual in being so attentive to the fine details of her *research*, while at the same time very much attuned to the broader aspects of her work and its medical implications. During her time in Livingood's group, she has become

> increasingly independent, and has been working on her own independently funded project for several years. At this point, she is clearly ready to establish her own *research* program.
> . . . She also writes well, and has authored several review articles in addition to her numerous *research* publications. As part of Livingood's large *research* group, Sarah has been a good team player who is sincerely interested in helping others in the group to succeed. Her thoughtfulness and caring for others certainly added greatly to the atmosphere of the lab while I was there.
> . . . I am certain that Sarah will be an extremely productive and valuable member of a university or medical school *research* department, and I highly recommend her to you.

A concern with research described as contribution to the broader research environment is that it may become a form of service, well appreciated but rarely grounds for promotion.

As with the status term 'research', there was also a contrast in the numbers of multiple text lines that contained *scientific terminology* between letters for women and letters for men. (We defined scientific terminology as words and phrases that go beyond reference to well-known organs and which are not found in standard dictionaries.) Scientific terminology can be another way of referring to research, but at the same time it is a broader category as it can be used to refer to clinical work or even teaching. In letters of recommendation it can be seen as a collusive feature, for the recommender knows that the evaluator, unlike the rest of society, understands such language.

A quarter of the letters for women used some form of scientific terminology, whereas a third of the men's letters used it. In terms of repetition of text units, that is lines of a letter with scientific terminology, here for the women's letters it was 1.9 text lines per letter, whereas for the men it was 3.3 lines per letter. Some would not call this repetition in that it is not necessarily exact words or phrases repeated. Pragmatically, however, use of scientific terminology functions both as a sign of knowledge of detail on the part of the recommender, as well as support for the significance of the applicant's work in that it warrants extended description in this restricted form.

Finally, another new category, keyed by repetition, evolved from experimental playing with concordance programs and their results. Not surprisingly, we found that 'her' and 'his' were high on the list of often-used words in the letters of recommendation. And we found that the overall usage of *possessive phrases referring to the applicant*, both pronominal and nominal as in 'his career', 'William's career'; 'her teaching', 'Sarah's teaching', was similar for letters for men and for women at 1.6 occurrences for every 10 lines of text. Further, there is a sense in which the presuppositional phrase, 'his research', is more closely associated with him than 'the research he conducts'. As we worked with the objects of these possessives, interesting patterns began to emerge.

But first, in surveying such third-person possessive phrases referring to the applicants, we felt the need to limit the data in two ways. First, we wanted equal

numbers of letters, so given that there were 89 letters for females, we took a random 89 letters for males in the database. Second, we wanted similar numbers of higher status letters. We defined these as of similar rank. Therefore, the letters for the full professor, who is male, were excluded, as were letters for associate rank beyond the numbers of such letters for female applicants. Then, after making lists of exact possessive phrases with their objects, we grouped these into semantic categories. Because the 89 letters for female applicants were about 5/6 or 84 percent as long as the random 89 letters for male applicants, we focus more on differential rank-ordering of semantic categories within genders than on contrasts in numbers of examples across genders, although we note these as well on the left side of Figures 3 and 4.

The most common semantic categories of objects of possessive phrases for female applicants were: 'her training', 'her teaching', and 'her application' (for the position). In contrast, for male applicants the most common semantic categories of objects of possessive phrases were: 'his research', 'his skills and abilities', and 'his career'. By this measure, the women are portrayed more as students and teachers, whereas the men are portrayed more as researchers and professionals. At a conference in Rhode Island when I (FT) presented this, a woman rose from her seat and said loudly, 'Yes'. It turned out she was a physician; through her efforts we were able to gain access to a second set of letters of recommendation for future study.

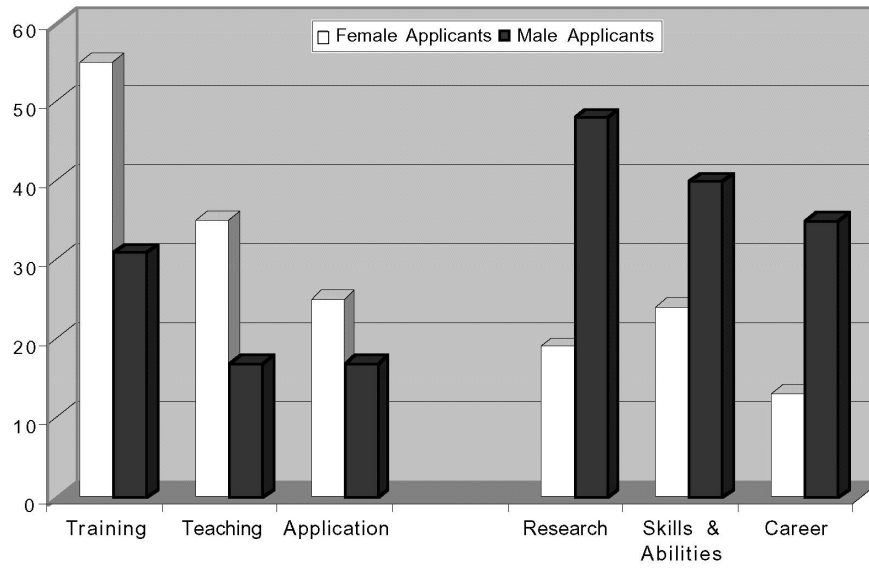


FIGURE 3. *Semantic realms following possessives. Rank-ordered within gender sets from equal numbers of letters 'her training'; 'his research'*

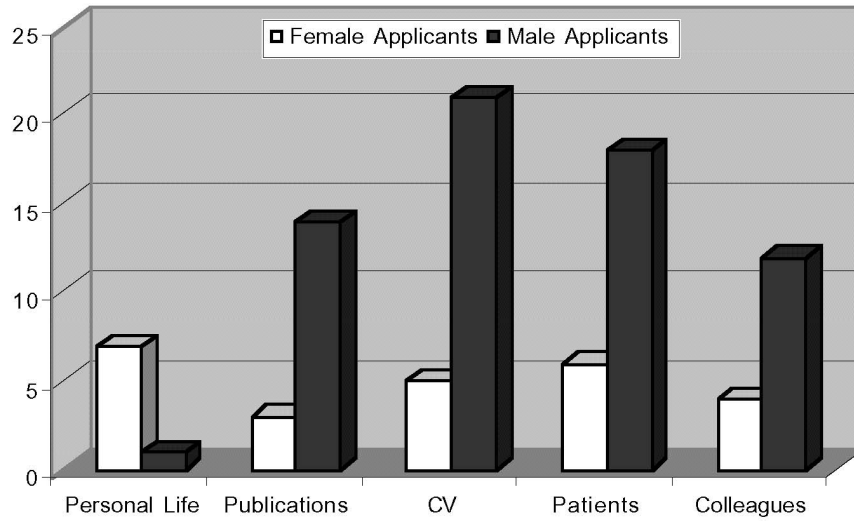


FIGURE 4. *Distinctive semantic realms following possessives. Greatest contrasts across genders in equal number of letters 'her personal life'; 'his publications'*

There is also another contrast that we were able to develop using the possessive phrases grouped by semantic realm. Here we sought the semantic realms that were most different by gender, keeping in mind that the women's letters were 84 percent the length of the men's.

Because the letters for men are 16 percent longer, there should be more references to 'personal life' than there are to women's, were this not a gendered realm. The number of references to 'his publications' appears enough greater to suggest that this is definitely more characteristic of letters for men. Could the much higher relative number of references to 'his CV' (curriculum vitae or résumé) suggest that the recommenders actually looked at the male applicants' CVs more? The greater number of references to 'his patients' is intriguing. Perhaps reference to women's 'interpersonal relations' and 'clinical skills' encompasses patients as well. The specificity here, however, is more compelling for males. Finally, the greater number of adjusted mentions of 'his colleagues' is something we want to research further. A generalization could be that the men are portrayed more as professional colleagues, whereas the women are seen more as students and teachers. When thinking of the potential for future advancement this is disquieting.

Examples of letters of female and male medical faculty by research directors

As a way of summarizing some of the many facets of this research, consider the following letters, both for clinical positions written by recommenders who were research directors of the applicants.

TEXT 14. EXAMPLE OF LETTER OF RECOMMENDATION FOR MALE APPLICANT

Dear Dr Koop:

William Harvey MD, has been a Postdoctoral Scholar in Pediatric Orthopedic Oncology at Northsouthern School of Medicine. During his time at our institution Bill actively *participated* in our research and educational activities. During his last year at Northsouthern he also *trained* in Pediatric Orthopedic Surgery.

Dr Harvey's research activities focused on the use of three dimensional contrast CT scanning for the assessment and quantification of blood flow and metabolism in solid bone tumors in the long bones of pediatric patients. An early study established a correlation between patterns of blood flow and tumor metabolism and the long term outcome of patients with solid tumors of long bones, while a second study . . . [+9 lines]
Some of this work has already been published in first rate orthopedic journals while other parts are currently under review by journals.

Bill's accomplishments are important for they demonstrate and underscore the clinical significance of altered patterns of blood flow and tumor metabolism for patient mortality and morbidity as well as for defining their implications for the management of patients with expected very poor prognosis for limb salvage. *His accomplishments* have been recognized locally by having been awarded two consecutive grants by the Greater Affiliate of the American Society for Orthopedic Surgeons. We believe it is also fair to state that *his accomplishments* have received, at least to some degree, national recognition as evidenced by several job offers.

Overall, we have found Bill to be a highly intelligent and hard working young man. He *communicates and collaborates* well with his peers and supervisors. On a more personal side, it saddened us to see him leave our institution yet we were not able to retain him for lack of funds. We believe that Bill is a promising, highly productive and creative young researcher who undoubtedly will become an independent and innovative investigator. Therefore, it is with considerable enthusiasm that we support unequivocally the proposed appointment to Assistant Professor of Pediatric Orthopedic Surgery and Oncology.

Sincerely,
Charles Lewis, MD
Chief, Dept. of Pediatric Oncology

In the above letter for William Harvey (Text 13), notice that what is not necessarily a title is listed as such, 'Postdoctoral Scholar in Pediatric Orthopedic Oncology', and mention is made of the rank for which he is applying, 'Assistant Professor'. The body of the letter is chock full of specifics of focus and record, including 13 lines of description of his research, some of which we omitted because of space considerations. In the third paragraph note the three repetitions of 'his accomplishments'. He is portrayed as agent in: 'Bill actively participated', 'he also trained', 'He communicates and collaborates well'. It could also be argued that 'Dr Harvey's research work focused on', and 'an early study established', show sociological agency if not linguistic agency (van Leeuwen, 1996: 32). The only negative refers to the recommender's own institution's inability to retain him.

TEXT 15. EXAMPLE OF LETTER OF RECOMMENDATION FOR FEMALE APPLICANT

Dear Alfred:

I am writing to you a letter of recommendation for *my good friend*, Dr Sarah Gray MD. As you probably know, I've known Sarah for about 7 years. I watched her career development while working at Northsouthern University, her presentations and prize winning events at the Academy of Pediatrics while a resident at Northsouthern and then her fellowship year with myself and Dr Dolittle in St. Louis some years ago.

Without any doubt, I am struck with Sarah's integrity. She is *totally intolerant of shoddy research work* and any work which has a hint of padding or error. Additionally, while working with her in St. Louis, I was able to watch her surgical skills. I felt she had been very well trained surgically in St. Louis but she has a *slight touch of lack of confidence* at times which I feel Sam Livingood is well aware of and will carefully work with Sarah regarding any matters like that during her clinical practice at Centvingtcinq.

I feel the addition of Sarah to the faculty of Centvingtcinq University and particularly to the Department of Cardiology of Children's Hospital to be a tremendous plus for that center. Her research work over the last few years has been 'top drawer' and virtually unchallengeable. I can only predict a great future for *this lady* and I am delighted that she has returned to further her career.

If you have any further questions about Sarah I'd be happy to discuss it with you.

Sincerely yours,
Charles Lewis, MD
Chief, Division of Cardiology

In the letter for Sarah Gray, notice the ambiguities of 'my friend', and 'she is totally intolerant of shoddy research'. And there is a singular lack of specificity in regard to her research. What is her research on? There is also paternalism and negative language in her 'slight touch of lack of confidence'. This can be seen as a way of bonding of recommender with the gatekeeper that works to the detriment of the applicant. The gendered naming is awkward – 'I can only predict a great future for this lady.' Why not 'for this cardiologist?' And unlike Bill, Sarah has a previous title, 'Head of Pediatric Cardiology', that is not mentioned in this letter, nor is there mention that she is applying for an Associate Professorship. Nowhere in this letter is she presented as the agent of her work.

Discussion

Which of the two above applicants is portrayed as more desirable as a colleague? Which is more likely to be promoted in the future? Which gives the better first impression? Recall that such letters of recommendation often arrive before the candidate steps in the door. And these letters stay in the file of the new faculty member. Further, if these letters are at all representative, it is not surprising that women receive fewer grants and advance at a much slower rate than men in

academic medicine. These letters and the findings of this research provoke many questions.

Briefly, in our extensive data we found that letters of recommendation for medical faculty differ in length in that a higher percentage of letters for women are very short (10% fewer than 10 lines), whereas a higher percentage of letters for men are very long (8% over 50 lines). Letters for female applicants are lacking in basic features to a statistically significant greater degree, and letters for women include *doubt raisers* at a statistically significant higher rate that is double the rate for males. We also found that there is a greater frequency of reference to terms of praise and the status category of 'research' in the letters of recommendation for men, as opposed to the letters for women. Finally, we found that when the possessive phrases relating to the applicant are grouped semantically, as a group the women's refer most to 'her teaching', 'her training', and 'her application'; whereas those for men as a group refer most to 'his research', 'his skills and abilities', and 'his career'.

Do these findings accurately reflect the applicants? Are the women less prepared? Have they done less research? Is medical training so stressful for women (Bowman and Allen, 1990) that they stay 'below the radar' to try to avoid harassment by teachers and fellow students (Tannen, 1994: 258), and competition with fellow students, and thus have fewer accomplishments and titles to be reported? Or are the female applicants of this data hired by the medical school precisely because they do not threaten the largely male gatekeepers?

Or do these findings reflect more the assumptions and prejudices of the largely male recommenders who are not used to women in potentially comparable high status positions? Do the largely male recommenders have little invested in the female applicants and therefore fail to exert themselves to write more complete letters?

Or as research in social and cognitive psychology has shown, have the recommenders merely fallen back on common societal gender schema in which women are not expected to have extensive accomplishments or even abilities in competitive professional work. That is, recommenders have unknowingly used selective categorization and perception, also known as stereotyping, in choosing what features to include in their profiles of the female applicants. Research in social psychology has found that such reliance on gender schema is more likely the more hierarchical the organization (Fiske, 1987: 43), and certainly American medical colleges are highly hierarchical. Such reliance on gender schema instead of individual description is also more likely, according to research, when a group is a minority in the institution (Valian, 1998: 141). In such a situation, members of the group are perceived first as a minority – here as women – and only secondarily as their profession – academic physicians and researchers. Finally, research in cognitive psychology has also found that when people cannot devote full attention to a task, they tend to rely more on gender schema (Valian, 1998: 308). In our interviews with academic physicians as well as in published accounts, it is clear that academic physicians have multiple, competing, and often unpredictable demands on their time.

Or is there something skewing our findings? The great advantage of working with naturalistic data – all the letters of successful applicants for three years – is that none were excluded for unknown reasons. Indeed, we have a picture of hiring for this institution for this period. And we assuredly have a picture that resonates with several members of the Committee for Hiring and Promotion. But a constraint is that certain practices, like hiring more men in research positions, are not controlled for. It is even difficult to know for sure which positions are research and which are clinical from the letters. For good clinicians also do research, particularly in their training time, and as the Head of Personnel told us, there are 24 faculty classifications of positions and recommenders often do not note them correctly. But even in singling out what appear to be research positions with no clinical component, the median length of letter for females was 289 words, whereas the median length for males was 336 words.

In a future study we would like to study letters of equal numbers of clinical and research applicants at equal ranks. The problem here, however, is that with current hiring trends the likelihood of finding letters of recommendation for female applicants for full professor positions is slight. Even with assistant and associate professor positions, it will take several years to accumulate letters for similar numbers of successful female research applicants as can be collected for successful male research applicants in one year.

The main contribution of this study is systematic methods for analyzing letters of recommendation, particularly for identifying categories that may not be mentioned in a letter but whose absence or infrequency is telling. Here knowledge of the context of a field, of what is of high status, in this case 'research', is an obvious first step, for language needs to be understood as constructed in communities of practice (Eckert and McConnell-Ginet, 1992: 487). Consideration of frequency of reference to such terms can be seen as a form of persuasiveness. Indeed, much research on letters of recommendation has neglected the vital feature of persuasion. Further, the category of *letters of minimal assurance*, that is, letters lacking in mention of basic features, is a new category in research of letters of recommendation. The expansion of 'negative language' to the broader category of *doubt raisers*, that include negative language, hedges, potentially negative, unexplained, faint praise, and irrelevancies, gives greater precision to the field. The use of *doubt raisers* in *apparent commendation* links this research to Critical Discourse Analysis of racist and prejudicial texts. Finally, the grouping of semantic categories of objects in possessive phrases referring to applicants, *semantic purses* of a group, is a contribution to the field. It provides a broader composite image of groups of applicants, as well as differences between groups. Comparing lists of such categories allows a signal way of noting absence or infrequent reference to a category for one group, that may be identified by its presence and frequency in letters for another group.

In conclusion, the response to the question from the member of the Executive Committee for Hiring and Promotion that triggered this research is that the letters for successful female applicants were indeed systematically different from

those for successful male applicants. Affirmative action programs[6] need to be cognizant of this research and incorporate its findings in their workshops. On a personal level, we hope this study stimulates recommenders to edit their own letters for unwarranted signs of gender schema and omissions of essential topics. We also encourage applicants to provide recommenders with readily usable descriptions of their research. More generally, we hope that this study stimulates future systematic studies of gatekeeping practices, particularly those that realize that gatekeeping does not stop at the door, but continues throughout professional life.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of Teun van Dijk to this article, Suzanne S. Toce, MD Professor of Pediatrics, St Louis University, as well as the member of the Executive Committee for Hiring and Promotion of the Medical School, who first approached us, and made the data available to us, but who must remain unnamed. Gabrielle Singleton, computer consultant, has worked on this project since its inception. Her fine design sense is well reflected in the figures, her facility with various software saved countless hours, and her enthusiasm for the project is most gratefully acknowledged.

NOTES

1. This study of the peer-review process of Sweden's Medical Research Council, one of the main funding agencies for biomedical research in Sweden, is remarkable in several ways. It was conducted in Sweden, ranked by the United Nations as the leading country in the world with respect to equal opportunities for men and women, and it is the first analysis based on actual peer-review scores, although the researchers had to go to court to obtain them. Its findings show that female applicants had to be 2.5 times more productive in terms of comparable publications to be accorded the same scientific competence score as male applicants. Applicants affiliated with the peer-review committee also received higher scientific competence scores than those of the same gender with equal productivity. (The particular peer review was of 114 applicants: 52 women and 62 men, of whom 20 were awarded postdoctoral fellowships: 4 women and 16 men.) The study 'strongly suggested that peer reviewers cannot judge scientific merit independent of gender' (Wenneras and Wold, 1997: 341).
2. We gratefully acknowledge Amy Sheldon for calling our attention to this important work, *Why So Slow? The Advancement of Women*, by cognitive psychologist, Virginia Valian.
3. Gender schemas (by convention not schemata) are defined by Valian (1998: 2) as 'a set of implicit, or nonconscious hypotheses about sex differences that play a central role in shaping men's and women's professional lives . . . (they) affect our expectations of men and women, our evaluations of their work, and their performance as professionals.' Valian sees stereotypes as one sort of such schema, but prefers the term 'schema' as being more inclusive, and less necessarily negative, acknowledging that it is a natural human activity to make hypotheses in making sense of the world. She sees gender schema as similar to a belief that the earth is flat; its naturalness does not guarantee its truthfulness. Only thinking the earth is flat will not make it so, whereas the cultural elaboration of gender schemas can lead to the creation of real sex differences (Valian, 1998: 118). Three processes work to entrench gender schemas in our minds: their

responsiveness to physical differences, our tendency to reason from extremes, and our tendency to see the sexes as dichotomous and gender traits as mutually exclusive (Valian, 1998: 119). Where Stokoe and Smithson (2001: 239) argued that conversational analysts tend to ignore culture and common-sense knowledge, we would add that this also involves ignoring gender schemas.

4. To preserve anonymity, the medical school to which the applicants were applying is referred to as 'Centvingtcinq University' as there were 125 medical schools in the US at that time. Similarly, all letters were addressed to Dr Alfred Koop, a made-up name recalling the former Surgeon General; all female applicants are referred to as Sarah Gray, the first woman to graduate in surgery from the University of Edinburgh and, a great-great-aunt of Frances Trix; and all male applicants are referred to William Harvey, who first published on the circulation of the blood in 1628. The hospitals are called 'Mrahonod', that is 'do no harm' spelled backwards; the applicants all come from 'Northsouthern University'; all trained in St Louis, a centrally located American city; and all male recommenders are named 'Charles Lewis'. We have also altered medical specialties to guard anonymity.
5. The special knowledge of the problems of the applicant could be seen to raise the credibility of the recommender. Another, more obvious example of 'apparent commendation' was found in a letter for an African American male applicant. 'He was a very good student, although not the best I saw in my 29 years there.' Here it is the experience of the recommender, not the ability of the student that is actually touted. Thus 'apparent commendation' builds on van Dijk's conceptualization, in that 'semantic moves of apparent denial or concession may be used in a combined strategy of positive self-presentation and negative presentation of the Other' (1993: 35).
6. An affirmative action program for female faculty in medical schools, conducted at Johns Hopkins University in the early 1990s (Fried et al., 1996), worked almost exclusively with women, setting up monthly meetings so they would know better what was expected for promotion, and addressed mentoring, but neglected to involve men in understanding how gender schema would affect evaluators in reasoning and judging (Valian, 1999: 320).

REFERENCES

Association of American Medical Colleges (1996) *Association of American Medical Colleges Data Book: Statistical Information Related to Medical Education* (eds P. Jolly and D. Hudley). Washington, DC: AAMC; AAMC Faculty Roster System Historical File (1995).

Bell, S., Cole, C.S. and Floge, L. (1992) 'Letters of Recommendation in Academe: Do Women and Men Write Different Languages?', *The American Sociologist* Fall: 7–22.

Bouton, L.F. (1995) 'A Cross-Cultural Analysis of the Structure and Content of Letters of Reference', *Studies in Second Language Acquisition* 17: 122–44.

Bowman, M.A. and Allen, D.I. (1990) *Stress and Women Physicians*. New York: Springer.

Brownlee, S. and Pezzullo, E. (1992) 'A Cure for Sexism: Women Doctors Herald a Kinder, Gentler Way to Practice Medicine', *U.S. News & World Report* March 23: 86–87, 90.

Conley, F.K. (1998) *Walking Out On the Boys*. New York: Farrar, Straus and Giroux.

Cowan, G. and Kasen, J.H. (1984) 'Form of Reference: Sex Differences in Letters of Recommendation', *Journal of Personality & Social Psychology* 46(3): 636–45.

DeLisa, J.A., Jain, S.S. and Campagnolo, D.I. (1994) 'Factors Used by Physical Medicine and Rehabilitation Residency Training Directors to Select Their Residents', *American Journal of Physical Medicine & Rehabilitation* 73(3): 152–6.

Drew, P. and Sorjonen, M-L. (1997) 'Institutional Dialogue', in T.A. van Dijk, (ed.), *Discourse as Social Interaction*. London: Sage.

Eckert, P. and McConnell-Ginet, S. (1992) 'Think Practically and Look Locally: Language and Gender as Community-Based Practice', *Annual Review of Anthropology* 21: 461–90.

Eger, H. (1991) 'Writer Perception, Writer Projection: The Influence of Personality, Ideology, and Gender on Letters of Recommendation', unpublished dissertation, University of Illinois at Chicago.

Ferrara, K.W. (1994) *Therapeutic Ways with Words*. New York: Oxford University Press.

Fiske, S.T. Trial Testimony 3/28/85, Tr. 534–633, in the Joint Appendix of the Briefs to the Supreme Court for Appellee in Price Waterhouse V. Ann B. Hopkins, Docket No. 87-1167, October Term, 1987, 21–82.

Fried, L.P., Francomano, C.A., MacDonald, S.M., Wagner, E.M., Stokes, E.J., Carbone, K.M., Bias, W.B., Newman, M.M. and Stobo, J.D. (1996) 'Career Development for Women in Academic Medicine: Multiple Interventions in a Department of Medicine', *Journal of the American Medial Association* 276: 898–905.

Gayed, N.M. (1991) 'Residency Directors Assessments of Which Selection Criteria Best Predict the Performances of Foreign-Born Medical Graduates During Internal Medicine Residencies', *Academic Medicine* 66(11): 699–701.

Goffman, E. (1981) 'Radio Talk', in *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Greenburg, A.G., Doyle, J. and McClure, D.K. (1994) 'Letters of Recommendation for Surgical Residencies: What They Say and What They Mean', *Journal of Surgical Research* 56(2): 192–8.

Gross, J. (1991) 'A Woman's Resignation Touches a Nerve at Medical Schools', *The New York Times* 14 July, 10.

Hatcher, L.L. (1983) 'Letters of Recommendation: Validity and Contribution to Variance in Judgements About Applicants', unpublished dissertation, Bowling Green State University, Ohio.

Johnson, M., Elam, C., Edwards, J., Taylor, D., Heldberg, C., Hinkley, R. and Comeau, R. (1998) 'Predicting Performance and Satisfaction: Beyond the Crystal Ball', *Academic Medicine* 73(10, Suppl.): S41–3.

LaCroix, P.P. (1985) 'Sex in Recs', *The Journal of College Admission* Fall: 24–6.

Morrisett, L.N. (1935) *Letters of Recommendation: A Study of Letters of Recommendation as an Instrument in the Selection of Secondary School Teachers*. New York: Columbia University.

O'Halloran, B.A., Altmaier E., Smith, W. and Franken, E. (1993) 'Evaluation of Resident Applicants by Letters of Recommendation: A Comparison of Traditional and Behavior-Based Formats', *Investigative Radiology* 28: 274–7.

Precht, K. (1998) 'A Cross-cultural Comparison of Letters of Recommendation', *English for Special Purposes* 17(3): 241–65.

Rayber, G.A.B. (1985) 'Source of Letters of Recommendation and Subsequent Teaching Performance of First-Year Teachers', unpublished dissertation, University of Georgia, Athens.

Selvin, P. (1993) 'Jenny Harrison Finally Gets Tenure in Math at Berkeley', *Science* 21: 287.

Stokoe, E.H. and Smithson, J. (2001) 'Making Gender Relevant: Conversation Analysis and Gender Categories in Interaction', *Discourse & Society* 12(2): 217–44.

Tannen, D. (1989) *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. New York: Cambridge University Press.

Tannen, D. (1994) *Talking from 9 to 5*. New York: Avon Books.

Tommasi, G., Williams, K. and Nordstrom, C. (1998) 'Letters of Recommendation: What

Information Captures HR Professionals' Attention?', *Journal of Business and Psychology* 13(1): 5–18.
Trix, F. and Sankar, A. (1998) 'Women's Voices and Experiences of the Hill/Thomas Hearings', *American Anthropologist* 100(1): 32–40.
Valian, V. (1998) *Why so Slow? The Advancement of Women*. Cambridge, MA: MIT Press.
Van Dijk, T.A. (1993) *Elite Discourse and Racism*. London: Sage.
Van Dijk, T.A., Ting-Toomey, S., Smitherman, G. and Troutman, D. (1997) 'Discourse, Ethnicity, Culture, and Racism', in T.A. van Dijk (ed.) *Discourse as Social Interaction*. London: Sage.
Van Leeuwen, T. (1996) 'The Representation of Social Actors', in C.R. Caldas-Coulthard and M. Coulthard (eds) *Texts and Practices: Readings in Critical Discourse Analysis*. London: Routledge.
Watson, C. (1987) 'Sex-Linked Differences in Letters of Recommendation', *Women and Language* X(2): 26–8.
Wenneras, C. and Wold, A. (1997) 'Nepotism and Sexism in Peer-Review', *Nature* 387: 341–3.
Yourcenar, M. (1981) *Discours Prononcé dans la Séance Publique*. Paris. http://www.academie-rancise.fr/imortels/discours_reception/yourcenar.html

FRANCES TRIX is Associate Professor of Anthropology at Wayne State University in Detroit. Her research interests include: attunement in interaction, transmission of spiritual knowledge, and gender. She is currently working on a discourse biography of Albanian Bektashi leader, Baba Rexheb. ADDRESS: Department of Anthropology, Wayne State University, 906 W. Warren, Detroit, MI 48202, USA. [email: ftrix@umich.edu]

CAROLYN PSENKA is a doctoral candidate in Organizational Anthropology at Wayne State University. Her research interests include the inter-relationship among risk, technology and culture; theories of complex organization; and gender. ADDRESS: Department of Anthropology, Wayne State University, 906 W. Warren, Detroit, MI 48202, USA. [email: psenka@aol.com]

The impact of gender on the review of the curricula vitae of job applicants a...
Rhea E Steinpreis; Katie A Anders; Dawn Ritzke
Sex Roles; Oct 1999; 41, 7/8; Research Library
pg. 509

Sex Roles, Vol. 41, Nos. 7/8, 1999

The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study

Rhea E. Steinpreis,[1] Katie A. Anders, and Dawn Ritzke
University of Wisconsin-Milwaukee

The purpose of this study was to determine some of the factors that influence outside reviewers and search committee members when they are reviewing curricula vitae, particularly with respect to the gender of the name on the vitae. The participants in this study were 238 male and female academic psychologists who listed a university address in the 1997 Directory of the American Psychological Association. They were each sent one of four versions of a curriculum vitae (i.e., female job applicant, male job applicant, female tenure candidate, and male tenure candidate), along with a questionnaire and a self-addressed stamped envelope. All the curricula vitae actually came from a real-life scientist at two different stages in her career, but the names were changed to traditional male and female names. Although an exclusively between-groups design was used to avoid sparking gender-conscious responding, the results indicate that the participants were clearly able to distinguish between the qualifications of the job applicants versus the tenure candidates, as evidenced by suggesting higher starting salaries, increased likelihood of offering the tenure candidates a job, granting them tenure, and greater respect for their teaching, research, and service records. Both men and women were more likely to vote to hire a male job applicant than a female job applicant with an identical record. Similarly, both sexes reported that the male job applicant had done adequate teaching, research, and service experience compared to the female job applicant with an identical record. In contrast, when men and women examined the highly competitive curriculum vitae of the real-life scientist who had gotten early tenure, they were equally likely to tenure the male and female tenure candidates and there was no difference in their ratings of their teaching, research, and service

[1]To whom correspondence should be addressed at Department of PsychologY, 138 Garland Hall, 2441 East Hartford Avenue, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53211; e-mail: rhea@csd.uwm.edu

0360-0025/99/1000-0509$16.00/0 © 1999 Plenum Publishing Corporation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

experience. There was no significant main effect for the quality of the institution or professional rank on selectivity in hiring and tenuring decisions. The results of this study indicate a gender bias for both men and women in preference for male job applicants.

The proportion of women receiving advanced degrees varies enormously by field, but psychology is one of the sciences which graduates one of the highest percentages of women. Fifty-eight percent of PhDs awarded in psychology go to women (Alper, 1993). Unfortunately, academia in psychology is disproportionately a male endeavor. Female academicians often deny the existence of gender discrimination, despite evidence to the contrary. In one study, female faculty members reported little evidence of discrimination even though an examination of personnel practices at their institution indicated gender discrepancies in initial appointments, promotion, salary, and tenure (Liss, 1975). This is not surprising given that women are more likely to attribute their success to chance, physical attractiveness, or affirmative action policies working in their favor (Clance, 1985).

The existence of discriminatory barriers to advancement that face women in organizations, including academia, has been well established (Morrison & Von Glinow, 1990; Northcraft & Gutek, 1993). These barriers are likely to include isolation, lack of peer and administrative support, increased likelihood of having to balance child-care responsibilities, lower income than their male counterparts, and lower status in their institution. Although the total number of faculty has increased over the past 20 years, the proportion of women faculty has remained the same. Throughout the world, women leave their academic careers more often than their male colleagues (Rothblum, 1988), and this occurs even after women get tenure (Herbold, 1995). Even after adjustment for productivity factors, women are less likely to be associate or full professors than are men with the same number of years in the field (Sonnert & Holton, 1996; Tesch, Wood, Helwig, & Nettinger, 1995).

Women are more likely than their male counterparts to experience social isolation and lack of peer support in many institutions (Ibarra, 1993; Ohlott, Ruderman, & McCauley, 1994). One way to attenuate the effects of social isolation and lack of peer support is to seek out a mentor within the system. Research has shown that individuals with mentors receive more promotions, have higher incomes (Dreher & Ash, 1990), report more career satisfaction (Fagenson, 1989; Turban & Dougherty, 1994), and report more career mobility (Scandura, 1992). Unfortunately, women face greater barri-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ers to developing a mentoring relationship than their male counterparts, even though they derive equivalent benefits if they are able to acquire a mentor (Dreher & Ash, 1990; Fagenson, 1989; Turban & Dougherty, 1994). For example, women working in male-dominated organizations are more likely than their majority counterparts to be in cross-gender mentoring relationships (Ragins & McFarlin, 1990). Indeed, other researchers have pointed out that the few women who have reached high levels in the business world are an essential population to study because they can provide critical information for future generations of women coming up in the field (Ragins, Townsend, & Mattis, 1998).

Thus, given the social isolation and difficulty finding a mentor, logic would dictate that women would need to learn to advocate for themselves. Negotiation skills are recognized as important for academic success (Applegate & Williams, 1990). However, research has shown that there are numerous obstacles associated with self-advocacy for women in academic settings (Janoff-Bulman & Wade, 1996). Women who are self-promoting, assertive, or dominant in interactions are evaluated more negatively than women who behave in a stereotype-consistent fashion and more negatively than men who are equally self-promoting, assertive, or dominant (Costrich, Feinstein, Kidder, Maracek, & Pascale, 1975; Linehan & Siefert, 1983, Rudman, 1995). There is also empirical evidence that there are problems associated with a woman's self-advocacy in initial starting salaries. Even with promotions and merit equal to that of their male colleagues, female faculty and employees continue to experience the monetary impact of lower starting salaries (Hallock, 1994). One potential obstacle to evening out the discrepancy between the salaries for the genders is the saliency of gender itself. It is difficult to hide one's gender for even a short period of time, let alone do this while pursuing a tenure-track position or pursuing tenure itself. Research has shown that gender is more salient for women when the person deciding their livelihood is male (Eagly & Karau, 1991). Jannoff-Bulman and Wade (1996) argue that men risk nothing by asking for higher salary and benefits because appearing assertive and self-promoting is a typical male behavior. In contrast, there are risks associated with a woman's self-advocacy, including being less influential in group exercises (Ridgeway, 1982; Ridgeway & Diekema, 1992). Conversely, women who demonstrated tentative, less self-assured speech were more influential, even though they were actually perceived as less competent and knowledgeable (Carli, 1989, 1990; Wiley & Eskilson, 1985). These researchers' findings on group dynamics have implications for academic departments. Many crucial decisions are made in department meetings following group discussions, including determining raises, deciding tenure, admitting graduate students, making curriculum decisions, etc. Many women may be afraid to advocate for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

themselves out of fear of getting punished in future decisions for speaking out in the past.

Little systematic data exist on the hiring processes in academic psychology departments (Sheehan, McDevitt, & Ross, 1998). Central to the processes of hiring and tenuring is the peer review of a job applicant's or candidate's work. Peer review has been criticized on many grounds, including poor interrater reliability, lack of objectivity, and nepotism (Cole, Cole, & Simon, 1981; Ernst, Resche, & Uher, 1992; Wenneras & Wold, 1997). Both women and men rate the quality of men's work higher than that of women when they are aware of the sex of the person to be evaluated, but not when the same person's gender is unknown (O'Leary & Wallston, 1982). In addition, female faculty members rate themselves lower than do their colleagues in teaching ability, number of publications, and professional reputation. In contrast, male faculty members view themselves more favorably than do their colleagues on these identical criteria (Widom and Burke, 1978). Female professors matched in rank and teaching experience receive lower teaching evaluations from students, particularly from male students, yet male faculty members are more likely than female faculty members to exhibit such behaviors as willingness to cancel class. Teaching style does not influence evaluations of male faculty to the same extent (Rothblum, 1988).

In the present study, we were interested in determining whether subjects would be influenced by the gender of the name on a curriculum vitae (CV) in determining the person's hireability and tenurability. To our knowledge this is the first study of its kind. The closest empirical data we could find was gathered almost 30 years ago. Fidell (1970) sent 155 department heads 10 descriptions of hypothetical candidates for faculty positions. The descriptions of the candidates were identical in all aspects except for gender, as implied by first names. Department heads were significantly more likely to indicate that they would hire female candidates at the assistant professor level and male candidates at the associate professor level. In the present study, we used CVs coming from a real scientist to promote believability. In addition, we sent the questionnaire to potential external reviewers at all faculty levels because department heads are certainly not the only ones to make hiring and tenure decisions. However, as in the Fidell study, our subjects spanned all disciplines within psychology. We also sought to limit the extent to which respondents would give politically correct answers by using a completely between-groups design (i.e., the subjects were sent only one vitae and were thus answering questions about only one gender), and by not asking the subjects their gender, although we knew their gender by discretely modifying the titles on the questionnaires for male versus female subjects.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

METHOD

The participants in this experiment were 238 academicians in United States who listed a university address in the Directory of the American Psychological Association (1997). The participants were picked in block random fashion, by selecting the first female academician on every other odd-numbered page (e.g., pages 1, 5, 9, etc.) and the first male academician on the alternating odd numbered pages (e.g., pages 3, 7, 11, etc.) Each subject's gender was determined by their name, and for the 7 androgynous names we encountered we discretely telephoned their departments and asked if they were male or female. We decided randomly to select the participant's names, rather than selecting equal numbers of male and female academicians at each rank (e.g., full, associate, and assistant professors). We felt that our procedure would ensure that the demographics of our sample would match the national trends in department composition, search committees, and tenuring executive committees. After all, it is this composition of individuals who are largely making the hiring and tenuring decisions. Thus, it is likely that our participant pool matches national trends in academic institutions, which would be middle to upper middle class Whites. Also, we did not ask subjects whether they were department heads because we were afraid they would fear the loss of their anonymity. The procedure we used yielded 800 potential participants. Only 582 had listed verifiable or deliverable addresses. Thus, 582 questionnaires were sent out to potential participants. Participants were asked on the questionnaire if they recognized any of the names appearing on the vitae and we elected to eliminate the 14 participants who indicated that they did recognize names (i.e., only the candidate's name was fake and all other names appearing on the CV were real scientists working in the field) to eliminate undue influence through familiarity. We also eliminated 4 participants who did not indicate that their primary work setting was in an academic department, all of whom told us that they were at a university counseling center. The final participant pool consisted of 118 males and 120 females. In the final analysis 238 of those questionnaires were returned to us, for a response rate of 41%. This response rate is consistent with similar surveys of professionals (Wunder & Wynn, 1988; McNevin, Leichner, Harper, & McCrimmon, 1985).

In order to prevent participants from trying to give gender-conscious responses, they were NOT asked their gender or name. However, we knew the male participants because they were sent questionnaires with the heading CURRICULUM VITAE STUDY and female participants were sent questionnaires with the heading CV STUDY. In addition, the quality of the institution was also known by stamping the back of the questionnaire with one of five different coded stamps of vague meaning in relation to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the study (i.e., *Department of Psychology, Confidential,* etc.). The quality of the program each participant listed in their address was coded as either first, second, third, or fourth quartile (or uncodeable) according to the National Research Council (1995). Therefore, when respondents returned their questionnaires we already knew the participant's gender and the quality of their institution.

The questionnaires asked participants whether they would hire the applicant, tenure the applicant, and what starting salary they would offer the applicant. They were also asked if the applicant had adequate teaching, research, and service experience to be hired or tenured. Participants were also asked to rank order what factors influenced them most when reviewing the vitae (e.g., research topics studied, number of publications and poster presentations, quality of publication journals, extramural funding history, applicant's training prior to applying, teaching contributions, or service contributions.) The questionnaire also gathered demographic information. Participants were asked on the questionnaire if they recognized any of the names appearing on the vitae. They were asked where their primary work setting was. They were asked their rank, number of publications, the number of times they had served as an external reviewer and the number of times they served on search committees. Participants were asked how many vita they had reviewed in their lifetime, whether they worked in a PhD-granting institution, and what their primary work setting was. They were also asked to rank order the qualities they looked for in a colleague (e.g., ability to establish an independent research program, collaboration prospects with you or other faculty members, collegiality/personality factors, or fulfilling affirmative action requirements). They were asked whether they thought they could serve as an adequate mentor to the applicant/candidate and also whether they personally supported the continuation of the tenure system.

Two versions of the CVs of a real-life scientist at different stages in her career were used as the review materials in this study. The first vitae was the one she had actually used to get a tenure-track job right out of graduate school and the second vitae was the one she had actually used to get early tenure. The use of a real research record would promote believability via real journal titles that are recognizable and empirical findings that are believable. Also, the real institutional affiliations were left on the vitae to promote believability. The real-life scientist was both a clinical psychologist and biological psychologist and this expertise was reflected in the sample vitae. Four sample CVs were used to represent the following four conditions: (1) a female job applicant, (2) a male job applicant, (3) a female tenure candidate, and (4) a male tenure candidate. The vitae included standard information on the scientist's educational background, current institutional affiliation, teaching, research, and service. The female and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

male job vitae at both levels were identical to each other, except for the use of a female or male name. In all cases, the female name used was "*Karen Miller*" and the male name used was "*Brian Miller.*" These names were selected because (1) they have been empirically demonstrated to be representative of each gender exclusively, without indication of either age or race (Kasof, 1993), and (2) to avoid confusing them with real psychologists, given there were no psychologists with those exact names appearing in the APA Membership Directory. In addition to changing the name of the applicant or candidate's CV, two other minor changes were made to the real-life CV: (1) Four years were added to each date appearing on the job applicant's CV, but not the tenure candidate's CV, in order to avoid giving the impression that the job applicant had been unemployed for several years, or had been working somewhere else and was denied tenure at that institution for publishing nothing, and (2) memberships in scientific groups for women were removed from all of the CVs, regardless of the name appearing on the CV, to avoid inducing subjects to hire or tenure the person because of any political ideology they may appear to have. Otherwise, the CVs were identical to the one the real-life scientist had used as a job applicant and a tenure candidate. The numerical contents of the CV sections of the job applicant and the tenure candidate are presented in Table I. The questionnaires accompanying each of these CVs were printed on different colored paper so that we could tell which group each subject was in, without having to ask the subjects whether they had received a male or female vitae. Using different-colored paper and different headings

Table I. The Numerical Values of the Curricula Vitae Used as Stimuli in this Study

Criterion	Job Applicant	Tenure Candidate
Formal education	Bachelor of Arts, Masters in Science, PhD in Psychology	Bachelor of Arts, Masters in Science, PhD in Psychology
Postdoctoral experience	None	None
Honors and awards	10	14
Publications	9	19
Manuscripts in submission	None	2
Manuscripts in preparation	None	3
Professional presentations	9	37
Invited presentations	None	4
Manuscripts review experience	None	3 Journals, 2 textbooks
Courses taught	2	10
Department committees	None	9
University committees	None	16
Community committees	None	2
Professional memberships	3	10

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

allowed us to avoid directly asking questions about gender, facilitating the disguise of purpose of the study.

Procedure

This study was approved by the investigator's Institutional Review Board for the Protection of Human Subjects. The participants were sent the questionnaire, a self-addressed envelope, a brief cover letter, and one of four different CVs. The cover letter informed participants that the purpose of the study was to examine the factors that influence the review of CVs during hiring and tenuring decisions. Participants were also told that their responses to the questionnaires were completely anonymous and that they should be frank in their responses.

RESULTS

The demographic information was analyzed on the pooled data, regardless of which curriculum vitae they had been sent. There were several significant differences between the male and female subjects which is consistent with national trends in psychology departments. Some of these differences are highlighted in Table II.

The male participants held higher academic ranks than the female participants, so they were significantly more likely to have tenure than the females ($t = 4.7$, $p < .0001$). The males also had more publications than the females ($t = 5.3$, $p < .0001$). The males had significantly more experience serving as an external reviewer for tenure cases ($t = 3.75$, $p > .0001$), and more experience serving on a search committee to hire a new faculty

Table II. Demographic Data of the Subject Pool by Subject Gender

Information category	Female Subjects	Male Subjects	Significance Value
Percent tenured	61%	87%	$p < .001$
Percent working in a PhD program	55%	68%	$p < .05$
Percent experienced as external reviewer on tenure cases	27%	41%	$p < .0001$
Biopsychology/neuroscience	4%	5%	$p > .05$
Clinical psychology	15%	18%	$p > .05$
Developmental psychology	12%	8%	$p > .05$
Experimental psychology	3%	12%	$p > .05$
Social/personality psychology	16%	6%	$p > .05$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

member ($t = 3.41$, $p < .001$). The males also reported examining more CVs in their lifetime than the females ($t = 2.62$, $p < .01$). The males were also more likely to work in a PhD-granting program than the females ($t = 2.37$, $p < .05$). Male and female participants did not differ in terms of their primary work setting ($t = -.24$, $p > .05$), which was predominantly in a university setting for both males and females. Male and female participants did not differ in terms of the quality of the program they were in, as indexed by the National Research Council's (1995) rating system ($t = .13$, $p > .05$). Also, male and female participants did not differ in the extent to which they reported being able to serve as an effective mentor to the person whose CV they were reviewing ($t = -.73$, $p > .05$). Participants were also asked if they supported the continuation of the tenure system, and the overwhelming majority of both male and female participants indicated that they did, with no significant gender difference in this trend ($t = -.75$, $p > .05$).

There were other similarities between the male and female participants. Males and females had a strikingly similar pattern of ranking the qualities that are most important to them in the selection of a new colleague. The participant's rankings of these qualities were analyzed using paired-samples t tests and the ranks presented below are significantly different from one another ($p < .001$). Both genders ranked "ability to establish an independent research program" as the most important quality. Both genders ranked "collegiality/personality factors" as the second most important quality, and "collaboration prospects with you or other faculty members" was ranked third. Interestingly, both genders ranked "fulfilling affirmative action requirements" last, with only five subjects in the entire pool ranking this quality as the most important.

Recall that participants were sent identical questionnaires regardless of whether they were sent the curriculum vitae of a hireable applicant or a tenure candidate. The data for the participant's responses regarding the hireability of all of the job applicants and tenure candidates are presented in Fig. 1 and their responses regarding their tenurability are presented in Fig. 2. These data were analyzed separately with respect to the questions being asked (e.g., hireability vs. tenurability), as well as the professional level of the target vitae (e.g., job applicants vs. tenure candidates). However, the data are combined in both Figs. 1 and 2 to illustrate clearly that the participants were in agreement about the general quality level of the CVs they received. Thus, even though we assumed that an individual with only nine publications would not ordinarily be given tenure, the participants were asked whether or not they would vote to give the job applicant *tenure*. This was done in an effort to make sure that our CV stimuli matched most academicians' perceptions of what is an appropriate record for hiring versus

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

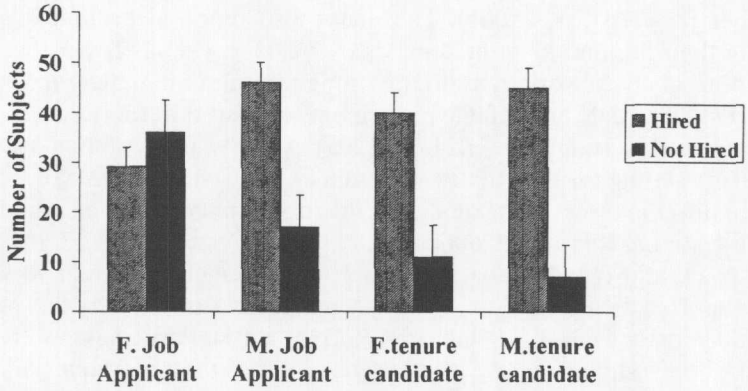


Fig. 1. Hireability of the job applicants and tenure candidates based on the quality of the curriculum vitae the participants were asked to evaluate.

tenuring. There were no significant main effects or interactions for the participants responding to the questions about tenuring an applicant who only had a hireable record, with the vast majority of the participants reporting that they would *not* vote to tenure them, and that they had not done enough teaching, research, and service.

As a further check on our hireable versus tenurable CV manipulation, we analyzed the data to determine if the participants selected a higher starting salary for the tenure candidates compared to the job applicants. An independent-samples *t* test indicated that all of the tenure candidates would have been offered a higher starting salary on average, compared to

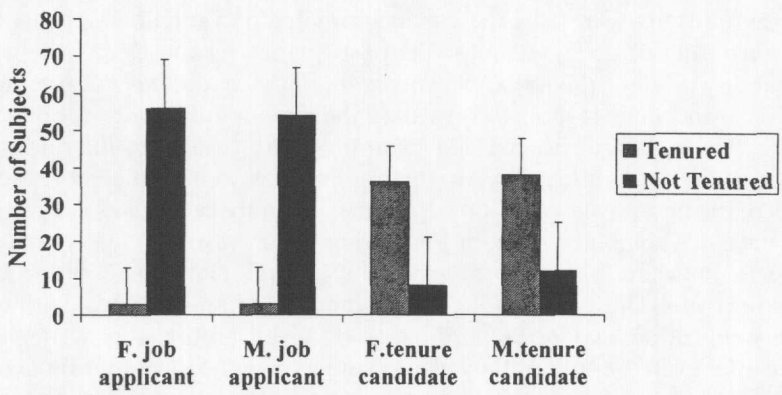


Fig. 2. Tenurability of the job applicants and tenure candidates based on the quality of the curriculum vitae the participants were asked to evaluate.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

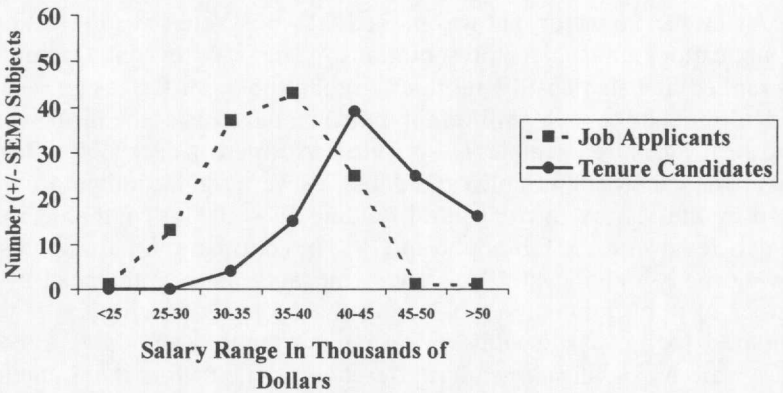


Fig. 3. The suggested starting salaries participants offered for the job applicants and tenure candidates.

the job applicants ($t = 11.69, p < .0001$). This was true regardless of whether or not the subjects voted to hire or tenure the applicant/candidate. These data are presented in Fig. 3. A 2 × 2 analysis of variance revealed no significant effects for the gender of the applicant, $F(1, 117) = .63, p > .05$, the gender of the participant, $F(1, 117) = .62, p > .05$, or the interaction between these two factors, $F(1, 117) = .17, p > .05$. Thus, there was no effect for gender on the selection of starting salaries.

We were also interested in whether or not participants were influenced by different factors when reviewing the job applicant versus tenure candidate CVs, and these results are shown in Fig. 4. Regardless of the CV they

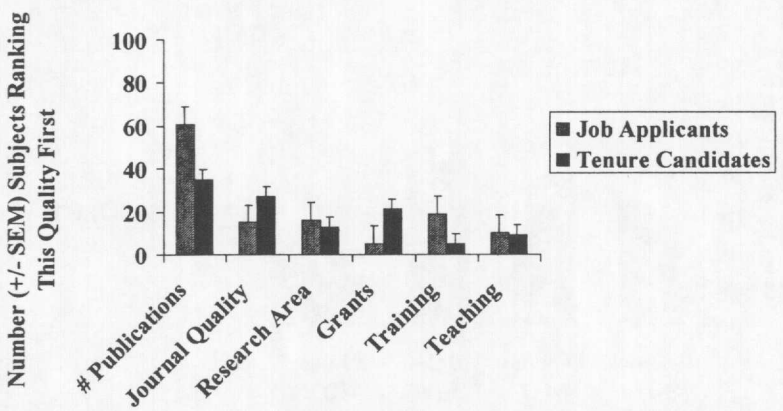


Fig. 4. The number of participants ranking each factor the most most influential in the review of the vitae of the job applicants versus the tenure candidates.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

had reviewed, the factor that was ranked first most often was the "number of publications and poster presentations." This was the only factor that was ranked first significantly more often than the other factors ($p < .001$). To determine if there were different influencing factors at the hiring versus tenuring level, paired-samples t tests were performed on each factor. Participants reviewing the tenurable candidate CVs were significantly more influenced by the history of extramural funding ($p < .0001$) compared to the subjects reviewing the job applicant CVs. In contrast, the participants reviewing the job applicant CVs were significantly more influenced by the number of publications ($p > .05$), and the subject's training ($p > .001$), compared to the participants reviewing the tenure candidates' records. There were no significant gender differences in the factors that influenced the participants when reviewing the CVs.

All of the remaining data were divided into groups of participants who had received hireable vita versus tenurable vita and these data sets were analyzed separately, using a 2 × 2 analysis of variance with a completely between-groups design. Participants were asked if they would vote to hire the person whose CV they were sent to review for a tenure track position in their department, and these results are shown in Fig. 5. There was a significant main effect for applicant gender, $F(1, 124) = 11.34$, $p < .001$, such that participants were more likely to hire the male applicants than the female applicants. There was not a significant main effect for the participant's gender $F(1, 124) = 1.31$, $p > .05$, so both males and females demonstrated the same gender bias in favor of male applicants. In addition, the interaction between the participant's gender and the hypothetical candidate was not significant, $F(1, 124) = 0.01$, $p > .05$. Participants were asked if they thought that the hypothetical job applicant had adequate research

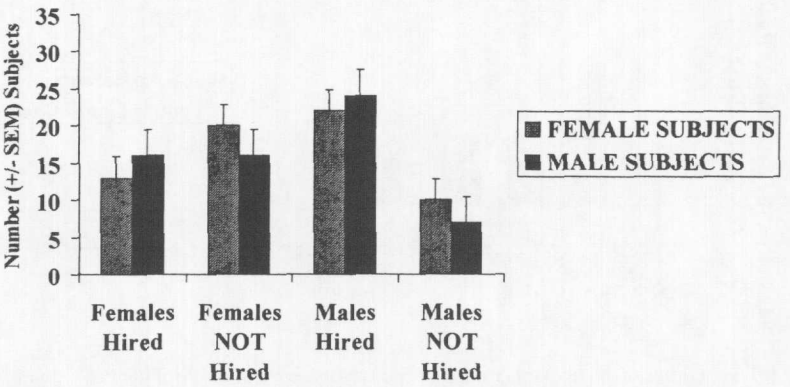


Fig. 5. Hireability of the job applicants as a function of the participant's gender.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

experience to be offered a tenure track position. There was a significant main effect for applicant gender, $F(1, 126) = 8.15, p < .005$, with participants being more likely to report that the male applicant had adequate research experience. There was no main effect for participant gender, $F(1, 126) = .88, p > .05$, and the interaction between the participant's gender and the hypothetical candidate was not significant, $F = 2.13, p > .05$. Participants were also asked if the applicant had adequate teaching and, separately, service experience. The data analysis yielded similar patterns to the first two questions, with significant main effects for the applicant's gender on adequacy of teaching experience, $F(1, 123) = 10.53, p < .005$, and service experience, $F(1, 119) = 8.97, p < .005$. In either case there was no main effect for the gender of the participant [$F(1, 123) = .37, p > .05$, and $F(1, 119) = .05, p > .05$, for teaching and service, respectively] or the interaction between gender of the participant and gender of the applicant [$F(1, 123) = .62, p > .05$, and $F(1, 119) = .99, p > .05$, for teaching and service, respectively].

A different pattern of results was found for the review of the tenure candidate's CVs, which was considerably more competitive for tenurability than the job applicant's CV was for getting hired. With respect to voting to tenure the candidate, there were no main effects for participant gender, $F(1, 102) = 1.44, p > .05$, candidate gender, $F(1, 102) = .07, p > .05$, or the interaction between these two factors, $F(1, 102) = 3.23, p > .05$. When participants were asked if the candidate had significant research experience, again there were no main effects for participant gender, $F(1, 103) = 1.42, p > .05$, candidate gender, $F(1, 103) = .1.42, p > .05$, or the interaction between these two factors, $F(1, 103) = 1.42, p > .05$. These results are reported in Fig. 6. There were no main effects for the gender of the partici-

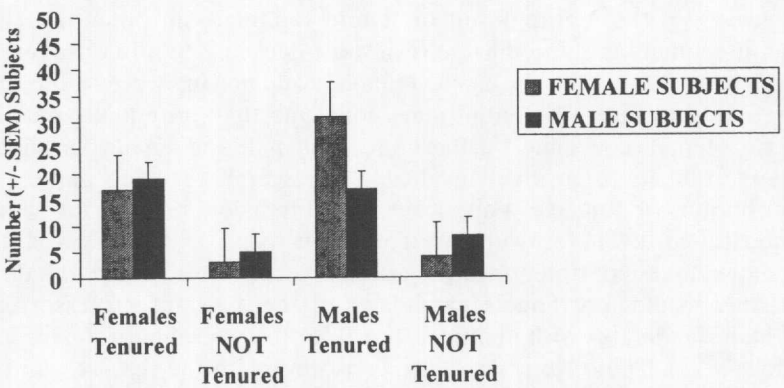


Fig. 6. Tenurability of the candidates by participant gender.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

pant, candidate gender, or the interaction between these two factors with respect to their perception of the adequacy of the candidate's teaching experience [$F(1, 99) = .7$, $p > .05$; $F(1, 99) = 1.2$, $p > .05$; and $F(1, 99) = .7$, $p > .05$, respectively]. There were also no main effects for the gender of the participant, candidate gender, or the interaction between these two factors with respect to their perception of the adequacy of the candidate's service contributions [$F(1, 102) = .14$, $p > .05$; $F(1, 102) = 1.35$, $p > .05$; and $F(1, 102) = .23$, $p > .05$, respectively].

Given that the sample vitae belonged to a biopsychologist, we did look to see if participants who shared the applicant/candidate's subdiscipline within psychology were more or less likely to hire or tenure the applicant. We found that the participant's subdiscipline made no difference in terms of whether or not they were likely to hire the job applicant, $F(6, 120) = .57$, $p > .05$, and the same was true of subdiscipline impact on the review of the tenure candidates, $F(6, 89) = 1.19$, $p > .05$). We did look to see if professional rank had an influence over whether or not participants were more or less likely to vote to hire or tenure the applicant or candidate. A 2×2 analysis of variance failed to reveal significant differences in votes between professors of different ranks in terms of their voting to hire, $F(3, 228) = .203$, $p > .05$, or voting to tenure, $F(3, 209) = .66$, $p > .05$. The impact of program quality on likelihood to hire or tenure was also examined. Again, there was no effect for participant rank on decision to hire the applicant, $F(4, 229) = 1.1$, $p > .05$, or tenure the candidate, $F(4, 209) = 1.4$, $p > .05$.

DISCUSSION

To our knowledge, this is the first study to determine the impact of the gender of the job applicant or tenure candidate on potential search committee member's and outside reviewer's decisions to hire or tenure. In the present study, both male and female academicians were significantly more likely to hire a potential male colleague than an equally qualified potential female colleague. Furthermore, both male and female participants were more likely to positively evaluate the research, teaching, and service contributions of a male job applicant than a female job applicant with an identical record. These results are consistent with previous research that has shown that department heads were significantly more likely to indicate that they would hire female candidates at the assistant professor level and male candidates with identical records at the associate professor level (Fidell, 1970). These results are also consistent with the research on how both women and men evaluate their own work (Widom and Burke, 1978)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and the work of others (O'Leary & Wallston, 1982). Indeed others have argued that although most men and women sincerely hold egalitarian beliefs, those beliefs alone do not guarantee impartial evaluation of others (Valian, 1998). The findings from this study support that contention and underscore the notion that women are as capable of gender bias as men are. Furthermore, these findings are particularly disturbing coming from psychologists, who unlike scientists in many other disciplines, would have been exposed to research on gender bias through course work, colleagues, and colloquia.

The present findings did not indicate that potential female tenure candidates are evaluated more negatively than potential male tenure candidates, although participants were four times as likely to write cautionary comments in the margins of their questionnaire if they had reviewed a female tenure candidate than if they had reviewed the male tenure candidate. These cautionary comments include such comments as, "We would have to see her job talk," "It is impossible to make such a judgement without teaching evaluations," "I would need to see evidence that she had gotten these grants and publications on her own." Such cautionary comments on the male tenure candidate's vitae were quite rare.

There are two possible explanations for the differential findings between hireability and tenurability. The first possible explanation is that academicians are somehow immune to gender bias when making tenure decisions, but not hiring decisions. It is intuitively appealing that it might be more difficult to make a decision that would cause someone to lose their job they already have than to turn someone down for a job they have not yet landed. However, the task in the present study with respect to tenuring was more similar to the task of an external reviewer, rather than an executive committee member voting on tenuring a colleague. Similarly, the methodology for the hiring conditions was akin to the task of search committee members prior to face-to-face interviews of the job applicants. Both tasks required the participants to make a decision about someone they had never met, and the participants were able to make equivalent decisions about the male and female tenure candidates. Therefore, we think it was unlikely that discomfort about making a decision about a colleague's job retention played a role in these findings.

We offer a more plausible explanation for the differential results between the tenurable and hireable findings. We believe the results are a direct reflection of the quality of the tenurable CV. To promote plausibility, we chose the CV of a real-life scientist at two different stages in her career for both the male and female CVs. The CVs were identical to the one the real-life scientist had used to get hired right out of graduate school and, 5 years later, tenured. We understood that the hireable vitae could *not* get

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

an applicant hired at many universities around the country, given the lack of postdoctoral experience, as well as limited research, teaching, and service experience. Our belief is supported by research that shows that significant predictors for success in pursuing an academic job in psychology are, in fact, publications and research grants awarded (Ng, 1997). Unfortunately, we believe that we underestimated the quality of the real-life tenurable CV. The real-life scientist had gotten early tenure at her institution, and had extensive research experience (frequent publications and extramural grants), teaching experience (i.e., including developing multiple courses at both the undergraduate and graduate levels), and service experience (membership in multiple committees at the departmental, university, state, and national level). The vast majority of the participants said they would tenure the tenurable candidate regardless of the gender of the candidate. In other words, our tenurable candidate appears to be infinitely more tenurable than our hireable candidate was hireable. The participants may have been responding to a kind of ceiling effect for the quality of the tenurable candidates. Taken together, these findings indicate that a superb record may indeed function as a buffer for gender bias when making promotional decisions.

We asked the participants what salary they would offer the applicant or tenure candidate. Although we realized that these decisions are usually out of the hands of individual faculty members, we knew that such individuals might have input into evaluating their colleagues during merit raise exercises or might be called upon by administrative bodies to suggest starting salaries if they are serving on search committees. We also hoped that their salary selections would match national starting levels, which would further legitimize the extent to which the study was a reflection of real-life hiring, tenuring, and pay scale decisions. The results indicated that there were no main effects for participant gender or applicant gender for salary selection. This finding could be a function of the presentation of the choices participants were given to endorse. There was a $5,000 gap within each choice. Perhaps this gap was too large to detect subtle gender biases in salary decisions that might be demonstrable if we had used narrower gaps of $1,000 or if we had used a free-form response for this question. Another possible explanation could be that participants were aware that their hiring and tenuring decisions are completely unlinked to their own professional lives. In contrast, universities sometimes have to make salary adjustments to level the playing field for gender-inequity reasons, for academicians who have similar records but dissimilar salary rates, etc. Perhaps the understanding that the salaries that participants select for job applicants or tenure candidates can be tied to their own in the future caused the participants to make fair decisions about salary offers. After all, the salaries that the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

entire participant pool selected were consistent with national trends in academic salaries (Black & Holden, 1998). Clearly further investigation of this important issue is of paramount concern to all academic disciplines, including psychology.

With respect to participants' beliefs about whether they would serve as an effective mentor to the job applicants or the tenure candidates, we found no overall difference between the genders. Although there was a nonsignificant trend for the female candidates to rate themselves as a poor potential mentor to the applicant/candidate more often and the male candidates to rate themselves as an excellent mentor, this is likely a function of the distribution of the genders in the ranks. In other words, as expected, the full professors were more likely to rate themselves as an excellent, good, or average mentor than the associate or assistant professors. This makes intuitive sense and other authors have pointed out that experienced mentors at high ranks may have a great ability to help their proteges' careers than novice mentors at lower ranks (Ragins, 1999). The mentorship picture is likely to be different from the mentee's point of view. Ragins and Cotton (1991) found that women are more likely than men to report restricted access to mentors. Although both male and female academicians may see themselves as competent mentors, the proportionately fewer female academicians taken together with the perceived barriers to getting a mentor may more closely reflect the state of mentorship in academia. Indeed, the proportion of women psychology mentors has increased, but not as rapidly as the proportion of women among doctoral recipients (Willis and Diebold, 1997).

It was interesting to note that there were no gender differences in terms of the qualities men and women were looking for in a colleague. In contrast to the myth that female academicians put all of their effort into fulfilling affirmative action goals, both men and women ranked this quality last. This may be a function of the desirability of the other options. Both genders were looking for the ability to establish an independent research program, as well as collegiality and personality factors. Research has shown that affirmative action is seen as relatively positive by both males and females (Parker, Baltes, & Christiansen, 1997), but that females are significantly more likely than males to think that the standard affirmative action practices are a good idea (Ozawa, Crosby & Crosby, 1996). However, the failure to see a lot of emphasis on fulfilling affirmative action goals in the present study is consistent with our findings that neither men or women seemed willing to give the female job applicant an equal chance. It may be that people agree with the concept in principle, but have difficulty applying the concept to hiring decisions.

The present findings indicate that at the fledgling stages of the career

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of a young professional, gender is seen as an indicator of success. Although it is unclear at what point the burgeoning record begins to speak for itself regardless of gender, it does seem clear that there comes a time when a scientist's record becomes strong enough to outweigh the gender bias. Two main lines of future inquiry seem mandatory. First, we need to understand better the specific factors that tip the scales in terms of ensuring that a record is evaluated on its own merit rather than in light of the scientist's gender. And second, academic departments need to be educated about this gender bias and as a discipline we need to develop selection systems that attenuate the gender bias by evaluating candidates on easily identifiable objective criteria. Obviously, this last step is a pretty tall order. Research has shown that both the job applicant's demeanor and sex-role stereotyping can influence hiring decisions (Gallois, Callan, & Palmer, 1993), such that candidates who used an assertive communication style were clearly favored by interviewers over aggressive and nonassertive candidates. Appearance and gender have been shown to impact hiring recommendations, particularly for positions that require high achievement, shrewdness, and leadership (Zebrowitz, Tenenbaum, & Goldstein, 1991). For example, males and mature-faced applicants are perceived as shrewder and more dominant than female and baby-faced applicants, and they are consequently favored for jobs that require such qualities. Furthermore, physical appearance in the absence of resume information has been linked to stereotype-guided processing in hiring recommendations (Branscombe & Smith, 1990). Thus, when there is no written documentation to go on, appearance is even more important. It is unlikely that gender-blind hiring and tenuring can ever be a reality, particularly when reviewers eventually end up considering the applicant's demeanor at the job talk and during interview. However, more research needs to be done on the factors that can promote fairness in the hiring process, and faculty need to be educated about the existence of this bias.

REFERENCES

Alper, J. (1993). The pipeline is leaking women all the way along. *Science, 260*, 409–411.

American Psychological Association. (1997). *Directory of the American Psychological Association.* Washington, DC: Author.

Applegate, W. B., & Williams, M. E. (1990). Career development in academic medicine. *American Journal of Medicine, 88*, 263–267.

Black, M. M., & Holden, E. W. (1998). The impact of gender on productivity and satisfaction among medical school psychologists. *Journal of Clinical Psychology in Medical Settings, 5*, 117–131.

Branscombe, N. R., & Smith, E. R. (1990). Gender and racial stereotypes in impression formation and social decision-making processes. *Sex Roles, 22*, 627–647.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Carli, L. L. (1989). Gender differences in interaction style and influence. *Journal of Personality and Social Psychology, 56,* 565–576.
Carli, L. L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology, 59,* 941–951.
Clance, P. R. (1985). *The Imposter Phenomenon: When Success Makes You Feel Like A Fake.* New York: Bantom Books.
Cole, S., Cole, J. R. & Simon, G. A. (1981). Chance and consensus in peer review. *Science, 214,* 881–886.
Costrich, N., Feinstein, J., Kidder, L., Maracek, J., & Pascale, L. (1975). When stereotypes hurt: Three studies of penalties for sex-role reversals. *Journal of Experimental Social Psychology, 11,* 520–530.
Dreher, G. F., & Ash, R. A. (1990). A comparative study of mentoring among men and women in managerial, professional, and technical positions. *Journal of Applied Psychology, 75,* 539–546.
Eagly, A. H., & Karau, S. J. (1991). Gender and the emergence of leaders: A meta-analysis. *Journal of Personality and Social Psychology, 60,* 685–710.
Ernst, E., Resch, K. L., & Uher, E. M. (1992). Reviewer Bias. *Annals of Internal Medicine, 116,* 958.
Fagenson, E. A. (1989). The mentor advantage: Perceived career/job experiences of proteges vs. non-proteges. *Journal of Organizational Behavior, 10,* 309–320.
Fidell, L. S. (1970). Empirical verification of sex discrimination in hiring practices in psychology. *American Psychologist, 25,* 1094–1098.
Gallois, C., Callan, V.J., & Palmer, J. A. M. (1993). The influence of applicant communication style and interviewer characteristics on hiring decisions. *Journal of Applied Social Psychology, 22,* 1041–1060.
Hallock, P. (1994). Promoting diversity on campus: Thought to action. *Thought and Action, X,* 65–78.
Herbold, H. (1995). Women who leave: Why women professors are cutting their ties to academia. *The Monthly Forum On Women In Higher Education,* 25–29.
Ibarra, H. (1993). Personal networks of women and minorities in management: A conceptual framework. *Academy of Management Review, 18,* 56–87.
Janoff-Bulman, R., & Wade, M. B. (1996). The dilemma of self-advocacy for women: Another case of blaming the victim? *Journal of Social and Clinical Psychology, 15(2),* 143–152.
Kasof, J. (1993). Sex bias in the name of stimulus persons. *Psychological Bulletin, 113,* 140- 163.
Linehan, M. M., & Seifert, R. F. (1983). Sex and contextual differences in the appropriateness of assertive behavior. *Psychology of Women Quarterly, 8,* 79–88.
Liss, L. (1975). Why academic women do not revolt: Implications for affirmative action. *Sex Roles, 1,* 209–230.
McNevin, S.H., Leichner, P., Harper, D., & McCrimmon, E. (1985). Sex role ideology among health care professionals. *Psychiatric Journal of the University of Ottawa, 10,* 21–23.
Morrison, A. M., & Von Glinow, M. A. (1990). Women and minorities in management. *American Psychologist, 45,* 200–208.
National Research Council (1995). *Research-Doctorate Programs in the United States: Continuity and Change,* pp. 371–377.
Ng, C. F. (1997). Recruitment practices and job search for academic positions in psychology. *Canadian Psychology, 38,* 25–42.
Northcraft, G. B., & Gutek, B. A. (1993). Point-counterpoint: Discrimination against women—Going, going, gone or going but never gone? In E. A. Fagenson (Ed.), *Women In Management: Trends, issues and challenges in managerial diversity* (pp. 219–245). Newbury Park, CA: Sage.
Ohlott, P. J., Ruderman, M. N., & McCauley, C. D. (1994). Gender differences in managers' developmental job experiences. *Academy of Management Journal, 37,* 46–67.
O'Leary, V. E., & Wallston, B. S. (1982). Women, gender, and social psychology. *Review of Personality and Social Psychology, 2,* 9–43.
Ozawa, K., Crosby, M., & Crosby, F. (1996). Individuals and resistance to affirmative action:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A comparison of Japanese and American samples. *Journal of Applied Social Psychology, 26,* 1138–1152.

Parker, C. P., Baltes, B. B., & Christiansen, N. D. (1997). Support for affirmative action, justice perceptions, and work attitudes: A study of gender and racial-ethnic group differences. *Journal of Applied Psychology, 82,* 376–389.

Ragins, B. R. (1999). Where do we go from here, and how do we get there? Methodological issues in conducting research on diversity and mentoring relationships. In A. J., Murrell, F. J. Crosby & R. J. Ely (Eds). *Mentoring Dilemmas: Developmental Relationships Within Multicultural Organizations* (pp. 227–246). Mahwah, NJ, Lawrence Erlbaum.

Ragins, B. R., & Cotton, J. L. (1991). Easier said than done: Gender differences in perceived barriers to gaining a mentor. *Academy of Management Journal, 34,* 939–951.

Ragins, B. R. & McFarlin, D. (1990). Perception of mentor roles in cross-gender mentoring relationships. *Journal of Vocational Behavior, 37,* 321–339.

Ragins, B. R., Townsend, B. & Mattis, M. (1998). Gender gap in the executive suite: CEOs and female executives report on breaking the glass ceiling. *Academy of Management Executive, 12,* 28–42.

Ridgeway, C. L. (1982). Status in groups: The importance of motivation. *American Sociological Review, 47,* 175–188.

Ridgeway, C. L., & Diekema, D. (1992). In C. L. Ridgeway (Ed.), *Gender, interaction, and inequity* (pp. 157–180). New York: Springer-Verlag.

Rothblum, E. D. (1988). Leaving the ivory tower: Factors contributing to women's voluntary resignation from academia. *Frontiers, 2,* 14–17.

Rudman, L. A. (1995). *To be or not to be (self-promoting): Motivational influences on gender stereotyping.* Poster presented at the 7th Annual Meeting of the American Psychological Society, New York.

Scandura, T. A. (1992). Mentorship and career mobility: An empirical investigation. *Journal of Organizational Behavior, 13,* 169–174.

Sheehan, E. P., McDevitt, T. M., & Ross, H. C. (1998). Looking for a job as a psychology professor? Factors affecting applicant success. *Teaching of Psychology, 25,* 8–11.

Sonnert, G., & Holton, G. (1996). Career patterns of women and men in the sciences. *American Scientist, 84,* 63.

Tesch, B. J., Wood, H. M., Helwig, A. L., & Nattinger, A. B. (1995). Promotion of women physicians in academic medicine: Glass ceiling or sticky floor? *Journal of the American Medical Association, 273(13),* 1022–1025.

Turban, D. B., & Dougherty, T. W. (1994). Role of protégé personality in receipt of mentoring and career success. *Academy of Management Journal, 37,* 688–702.

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

Wenneras, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature, 387,* 341–343.

Widom, C. S., & Burke, B. W. (1978). Performance, attitudes, and professional socialization of women in academia. *Sex Roles, 4,* 549–562.

Wiley, M. G., & Eskilson, A. (1985). Speech style, gender stereotypes, and corporate success: What if women talk more like men? *Sex Roles, 12,* 993–1006.

Willis, F. N., & Diebold, C. T. (1997). Producing mentors in psychology. *Teaching of Psychology, 24,*15–21.

Wunder, G. C., & Wynn, G. W. (1988). The effects of address personalization on mailed questionnaires response rate, time, and quality. *Journal of the Market Research Society, 30,* 95–101.

Zebrowitz, L. A., Tenenbaum, D. R., & Goldstein, L. H. (1991). The impact of job applicant's facial maturity, gender, and academic achievement on hiring recommendations. *Journal of Applied Social Psychology, 21,* 525–548.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

By MARIANNE BERTRAND AND SENDHIL MULLAINATHAN*

We study race in the labor market by sending fictitious resumes to help-wanted ads in Boston and Chicago newspapers. To manipulate perceived race, resumes are randomly assigned African-American- or White-sounding names. White names receive 50 percent more callbacks for interviews. Callbacks are also more responsive to resume quality for White names than for African-American ones. The racial gap is uniform across occupation, industry, and employer size. We also find little evidence that employers are inferring social class from the names. Differential treatment by race still appears to still be prominent in the U.S. labor market. (*JEL* J71, J64).

Every measure of economic success reveals significant racial inequality in the U.S. labor market. Compared to Whites, African-Americans are twice as likely to be unemployed and earn nearly 25 percent less when they are employed (Council of Economic Advisers, 1998). This inequality has sparked a debate as to whether employers treat members of different races differentially. When faced with observably similar African-American and White applicants, do they favor the White one? Some argue yes, citing either employer prejudice or employer perception that race signals lower productivity. Others argue that differential treatment by race is a relic of the past, eliminated by some combination of employer enlightenment, affirmative action programs and the profit-maximization motive. In fact, many in this latter camp even feel that stringent enforcement of affirmative action programs has produced an environment of reverse discrimination. They would argue that faced with identical candidates, employers might favor the African-American one.[1] Data limitations make it difficult to empirically test these views. Since researchers possess far less data than employers do, White and African-American workers that appear similar to researchers may look very different to employers. So any racial difference in labor market outcomes could just as easily be attributed to differences that are observable to employers but unobservable to researchers.

To circumvent this difficulty, we conduct a field experiment that builds on the correspondence testing methodology that has been primarily used in the past to study minority outcomes in the United Kingdom.[2] We send resumes in response to help-wanted ads in Chicago and Boston newspapers and measure callback for interview for each sent resume. We

* Bertrand: Graduate School of Business, University of Chicago, 1101 E. 58th Street, R0 229D, Chicago, IL 60637, NBER, and CEPR (e-mail: marianne.bertrand@gsb.uchicago.edu); Mullainathan: Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, E52-380a, Cambridge, MA 02142, and NBER (e-mail: mullain@mit.edu). David Abrams, Victoria Bede, Simone Berkowitz, Hong Chung, Almudena Fernandez, Mary Anne Guediguian, Christine Jaw, Richa Maheswari, Beverley Martis, Alison Tisza, Grant Whitehorn, and Christine Yee provided excellent research assistance. We are also grateful to numerous colleagues and seminar participants for very helpful comments.

[1] This camp often explains the poor performance of African-Americans in terms of supply factors. If African-Americans lack many basic skills entering the labor market, then they will perform worse, even with parity or favoritism in hiring.

[2] See Roger Jowell and Patricia Prescott-Clarke (1970), Jim Hubbuck and Simon Carter (1980), Colin Brown and Pat Gay (1985), and Peter A. Riach and Judith Rich (1991). One caveat is that some of these studies fail to fully match skills between minority and nonminority resumes. For example some impose differential education background by racial origin. Doris Weichselbaumer (2003, 2004) studies the impact of sex-stereotypes and sexual orientation. Richard E. Nisbett and Dov Cohen (1996) perform a related field experiment to study how employers' response to a criminal past varies between the North and the South in the United States.

experimentally manipulate perception of race via the name of the fictitious job applicant. We randomly assign very White-sounding names (such as Emily Walsh or Greg Baker) to half the resumes and very African-American-sounding names (such as Lakisha Washington or Jamal Jones) to the other half. Because we are also interested in how credentials affect the racial gap in callback, we experimentally vary the quality of the resumes used in response to a given ad. Higher-quality applicants have on average a little more labor market experience and fewer holes in their employment history; they are also more likely to have an e-mail address, have completed some certification degree, possess foreign language skills, or have been awarded some honors.[3] In practice, we typically send four resumes in response to each ad: two higher-quality and two lower-quality ones. We randomly assign to one of the higher- and one of the lower-quality resumes an African-American-sounding name. In total, we respond to over 1,300 employment ads in the sales, administrative support, clerical, and customer services job categories and send nearly 5,000 resumes. The ads we respond to cover a large spectrum of job quality, from cashier work at retail establishments and clerical work in a mail room, to office and sales management positions.

We find large racial differences in callback rates.[4] Applicants with White names need to send about 10 resumes to get one callback whereas applicants with African-American names need to send about 15 resumes. This 50-percent gap in callback is statistically significant. A White name yields as many more callbacks as an additional eight years of experience on a resume. Since applicants' names are randomly assigned, this gap can only be attributed to the name manipulation.

Race also affects the reward to having a better resume. Whites with higher-quality resumes receive nearly 30-percent more callbacks than Whites with lower-quality resumes. On the other hand, having a higher-quality resume has a smaller effect for African-Americans. In other words, the gap between Whites and African-Americans widens with resume quality. While one may have expected improved credentials to alleviate employers' fear that African-American applicants are deficient in some unobservable skills, this is not the case in our data.[5]

The experiment also reveals several other aspects of the differential treatment by race. First, since we randomly assign applicants' postal addresses to the resumes, we can study the effect of neighborhood of residence on the likelihood of callback. We find that living in a wealthier (or more educated or Whiter) neighborhood increases callback rates. But, interestingly, African-Americans are not helped more than Whites by living in a "better" neighborhood. Second, the racial gap we measure in different industries does not appear correlated to Census-based measures of the racial gap in wages. The same is true for the racial gap we measure in different occupations. In fact, we find that the racial gaps in callback are statistically indistinguishable across all the occupation and industry categories covered in the experiment. Federal contractors, who are thought to be more severely constrained by affirmative action laws, do not treat the African-American resumes more preferentially; neither do larger employers or employers who explicitly state that they are "Equal Opportunity Employers." In Chicago, we find a slightly smaller racial gap when employers are located in more African-American neighborhoods.

The rest of the paper is organized as follows. Section I compares this experiment to earlier work on racial discrimination, and most notably to the labor market audit studies. We describe the experimental design in Section II and present the results in Section III, subsection A. In Section IV, we discuss possible interpretations of our results, focusing especially on two issues. First, we examine whether the

[3] In creating the higher-quality resumes, we deliberately make small changes in credentials so as to minimize the risk of overqualification.

[4] For ease of exposition, we refer to the effects uncovered in this experiment as racial differences. Technically, however, these effects are about the racial soundingness of names. We briefly discuss below the potential confounds between name and race. A more extensive discussion is offered in Section IV, subsection B.

[5] These results contrast with the view, mostly based on nonexperimental evidence, that African-Americans receive higher returns to skills. For example, estimating earnings regressions on several decades of Census data, James J. Heckman et al. (2001) show that African-Americans experience higher returns to a high school degree than Whites do.

race-specific names we have chosen might also proxy for social class above and beyond the race of the applicant. Using birth certificate data on mother's education for the different first names used in our sample, we find little relationship between social background and the name-specific callback rates.[6] Second, we discuss how our results map back to the different models of discrimination proposed in the economics literature. In doing so, we focus on two important results: the lower returns to credentials for African-Americans and the relative homogeneity of the racial gap across occupations and industries. We conclude that existing models do a poor job of explaining the full set of findings. Section V concludes.

I. Previous Research

With conventional labor force and household surveys, it is difficult to study whether differential treatment occurs in the labor market.[7] Armed only with survey data, researchers usually measure differential treatment by comparing the labor market performance of Whites and African-Americans (or men and women) for which they observe similar sets of skills. But such comparisons can be quite misleading. Standard labor force surveys do not contain all the characteristics that employers observe when hiring, promoting, or setting wages. So one can never be sure that the minority and nonminority workers being compared are truly similar from the employers' perspective. As a consequence, any measured differences in outcomes could be attributed to these unobserved (to the researcher) factors.

This difficulty with conventional data has led some authors to instead rely on pseudo-experiments.[8] Claudia Goldin and Cecilia Rouse (2000), for example, examine the effect of blind auditioning on the hiring process of orchestras. By observing the treatment of female candidates before and after the introduction of blind auditions, they try to measure the amount of sex discrimination. When such pseudo-experiments can be found, the resulting study can be very informative; but finding such experiments has proven to be extremely challenging.

A different set of studies, known as audit studies, attempts to place comparable minority and White actors into actual social and economic settings and measure how each group fares in these settings.[9] Labor market audit studies send comparable minority (African-American or Hispanic) and White auditors in for interviews and measure whether one is more likely to get the job than the other.[10] While the results vary somewhat across studies, minority auditors tend to perform worse on average: they are less likely to get called back for a second interview and, conditional on getting called back, less likely to get hired.

These audit studies provide some of the cleanest nonlaboratory evidence of differential treatment by race. But they also have weaknesses, most of which have been highlighted in Heckman and Siegelman (1992) and Heckman (1998). First, these studies require that both members of the auditor pair are identical in all dimensions that might affect productivity in employers' eyes, except for race. To accomplish this, researchers typically match auditors on several characteristics (height, weight, age, dialect, dressing style, hairdo) and train them for several days to coordinate interviewing styles. Yet, critics note that this is unlikely to erase the numerous differences that exist between the auditors in a pair.

Another weakness of the audit studies is that they are not double-blind. Auditors know the purpose of the study. As Turner et al. (1991)

[6] We also argue that a social class interpretation would find it hard to explain some of our findings, such as why living in a better neighborhood does not increase callback rates more for African-American names than for White names.

[7] See Joseph G. Altonji and Rebecca M. Blank (1999) for a detailed review of the existing literature on racial discrimination in the labor market.

[8] William A. Darity, Jr. and Patrick L. Mason (1998) describe an interesting nonexperimental study. Prior to the Civil Rights Act of 1964, employment ads would explicitly state racial biases, providing a direct measure of differential treatment. Of course, as Arrow (1998) mentions, discrimination was at that time "a fact too evident for detection."

[9] Michael Fix and Marjery A. Turner (1998) provide a survey of many such audit studies.

[10] Earlier hiring audit studies include Jerry M. Newman (1978) and Shelby J. McIntyre et al. (1980). Three more recent studies are Harry Cross et al. (1990), Franklin James and Steve W. DelCastillo (1991), and Turner et al. (1991). Heckman and Peter Siegelman (1992), Heckman (1998), and Altonji and Blank (1999) summarize these studies. See also David Neumark (1996) for a labor market audit study on gender discrimination.

note: "The first day of training also included an introduction to employment discrimination, equal employment opportunity, and a review of project design and methodology." This may generate conscious or subconscious motives among auditors to generate data consistent or inconsistent with their beliefs about race issues in America. As psychologists know very well, these demand effects can be quite strong. It is very difficult to insure that auditors will not want to do "a good job." Since they know the goal of the experiment, they can alter their behavior in front of employers to express (indirectly) their own views. Even a small belief by auditors that employers treat minorities differently can result in measured differences in treatment. This effect is further magnified by the fact that auditors are not in fact seeking jobs and are therefore more free to let their beliefs affect the interview process.

Finally, audit studies are extremely expensive, making it difficult to generate large enough samples to understand nuances and possible mitigating factors. Also, these budgetary constraints worsen the problem of mismatched auditor pairs. Cost considerations force the use of a limited number of pairs of auditors, meaning that any one mismatched pair can easily drive the results. In fact, these studies generally tend to find significant differences in outcomes across pairs.

Our study circumvents these problems. First, because we only rely on resumes and not people, we can be sure to generate comparability across race. In fact, since race is randomly assigned to each resume, the same resume will sometimes be associated with an African-American name and sometimes with a White name. This guarantees that any differences we find are caused solely by the race manipulation. Second, the use of paper resumes insulates us from demand effects. While the research assistants know the purpose of the study, our protocol allows little room for conscious or subconscious deviations from the set procedures. Moreover, we can objectively measure whether the randomization occurred as expected. This kind of objective measurement is impossible in the case of the previous audit studies. Finally, because of relatively low marginal cost, we can send out a large number of resumes. Besides giving us more precise estimates, this larger sample size also allows us to examine the nature of the differential treatment from many more angles.

II. Experimental Design

A. *Creating a Bank of Resumes*

The first step of the experimental design is to generate templates for the resumes to be sent. The challenge is to produce a set of realistic and representative resumes without using resumes that belong to actual job seekers. To achieve this goal, we start with resumes of actual job searchers but alter them sufficiently to create distinct resumes. The alterations maintain the structure and realism of the initial resumes without compromising their owners.

We begin with resumes posted on two job search Web sites as the basis for our artificial resumes.[11] While the resumes posted on these Web sites may not be completely representative of the average job seeker, they provide a practical approximation.[12] We restrict ourselves to people seeking employment in our experimental cities (Boston and Chicago). We also restrict ourselves to four occupational categories: sales, administrative support, clerical services, and customer services. Finally, we further restrict ourselves to resumes posted more than six months prior to the start of the experiment. We purge the selected resumes of the person's name and contact information.

During this process, we classify the resumes within each detailed occupational category into two groups: high and low quality. In judging resume quality, we use criteria such as labor market experience, career profile, existence of gaps in employment, and skills listed. Such a classification is admittedly subjective but it is made independently of any race assignment on the resumes (which occurs later in the experimental design). To further reinforce the quality gap between the two sets of resumes, we add to each high-quality resume a subset of the following features: summer or while-at-school employment experience, volunteering experience, extra computer skills, certification degrees, foreign language skills, honors, or some military

[11] The sites are www.careerbuilder.com and www.americasjobbank.com.

[12] In practice, we found large variation in skill levels among people posting their resumes on these sites.

experience. This resume quality manipulation needs to be somewhat subtle to avoid making a higher-quality job applicant overqualified for a given job. We try to avoid this problem by making sure that the features listed above are not all added at once to a given resume. This leaves us with a high-quality and a low-quality pool of resumes.[13]

To minimize similarity to actual job seekers, we use resumes from Boston job seekers to form templates for the resumes to be sent out in Chicago and use resumes from Chicago job seekers to form templates for the resumes to be sent out in Boston. To implement this migration, we alter the names of the schools and previous employers on the resumes. More specifically, for each Boston resume, we use the Chicago resumes to replace a Boston school with a Chicago school.[14] We also use the Chicago resumes to replace a Boston employer with a Chicago employer in the same industry. We use a similar procedure to migrate Chicago resumes to Boston.[15] This produces distinct but realistic looking resumes, similar in their education and career profiles to this subpopulation of job searchers.[16]

B. *Identities of Fictitious Applicants*

The next step is to generate identities for the fictitious job applicants: names, telephone numbers, postal addresses, and (possibly) e-mail addresses. The choice of names is crucial to our experiment.[17] To decide on which names are uniquely African-American and which are uniquely White, we use name frequency data calculated from birth certificates of all babies born in Massachusetts between 1974 and 1979. We tabulate these data by race to determine which names are distinctively White and which are distinctively African-American. Distinctive names are those that have the highest ratio of frequency in one racial group to frequency in the other racial group.

As a check of distinctiveness, we conducted a survey in various public areas in Chicago. Each respondent was asked to assess features of a person with a particular name, one of which is race. For each name, 30 respondents were asked to identify the name as either "White," "African-American," "Other," or "Cannot Tell." In general, the names led respondents to readily attribute the expected race for the person but there were a few exceptions and these names were disregarded.[18]

The final list of first names used for this study is shown in Appendix Table A1. The table reports the relative likelihood of the names for the Whites and African-Americans in the Massachusetts birth certificates data as well as the recognition rate in the field survey.[19] As Appendix Table A1 indicates, the African-American first names used in the experiment are quite common in the population. This suggests that by using these names as an indicator of race, we are actually covering a rather large segment of the African-American population.[20]

Applicants in each race/sex/city/resume quality cell are allocated the same phone number. This guarantees that we can precisely track employer callbacks in each of these cells. The phone lines we use are virtual ones with only a voice mailbox attached to them. A similar outgoing message is recorded on each of the voice mailboxes but each message is recorded by someone of the appropriate race and gender.

[13] In Section III, subsection B, and Table 3, we provide a detailed summary of resume characteristics by quality level.

[14] We try as much as possible to match high schools and colleges on quality and demographic characteristics.

[15] Note that for applicants with schooling or work experience outside of the Boston or Chicago areas, we leave the school or employer name unchanged.

[16] We also generate a set of different fonts, layouts, and cover letters to further differentiate the resumes. These are applied at the time the resumes are sent out.

[17] We chose name over other potential manipulations of race, such as affiliation with a minority group, because we felt such affiliations may especially convey more than race.

[18] For example, Maurice and Jerome are distinctively African-American names in a frequency sense yet are not perceived as such by many people.

[19] So many of names show a likelihood ratio of ∞ because there is censoring of the data at five births. If there are fewer than five babies in any race/name cell, it is censored (and we do not know whether a cell has zero or was censored). This is primarily a problem for the computation of how many African-American babies have "White" names.

[20] We also tried to use more White-sounding last names for White applicants and more African-American-sounding last names for African-American applicants. The last names used for White applicants are: Baker, Kelly, McCarthy, Murphy, Murray, O'Brien, Ryan, Sullivan, and Walsh. The last names used for African-American applicants are: Jackson, Jones, Robinson, Washington, and Williams.

Since we allocate the same phone number for applicants with different names, we cannot use a person name in the outgoing message.

While we do not expect positive feedback from an employer to take place via postal mail, resumes still need postal addresses. We therefore construct fictitious addresses based on real streets in Boston and Chicago using the White Pages. We select up to three addresses in each 5-digit zip code in Boston and Chicago. Within cities, we randomly assign addresses across all resumes. We also create eight e-mail addresses, four for Chicago and four for Boston.[21] These e-mail addresses are neutral with respect to both race and sex. Not all applicants are given an e-mail address. The e-mail addresses are used almost exclusively for the higher-quality resumes. This procedure leaves us with a bank of names, phone numbers, addresses, and e-mail addresses that we can assign to the template resumes when responding to the employment ads.

C. *Responding to Ads*

The experiment was carried out between July 2001 and January 2002 in Boston and between July 2001 and May 2002 in Chicago.[22] Over that period, we surveyed all employment ads in the Sunday editions of *The Boston Globe* and *The Chicago Tribune* in the sales, administrative support, and clerical and customer services sections. We eliminate any ad where applicants were asked to call or appear in person. In fact, most of the ads we surveyed in these job categories ask for applicants to fax in or (more rarely) mail in their resume. We log the name (when available) and contact information for each employer, along with any information on the position advertised and specific requirements (such as education, experience, or computer skills). We also record whether or not the ad explicitly states that the employer is an equal opportunity employer.

For each ad, we use the bank of resumes to sample four resumes (two high-quality and two low-quality) that fit the job description and requirements as closely as possible.[23] In some cases, we slightly alter the resumes to improve the quality of the match, such as by adding the knowledge of a specific software program.

One of the high- and one of the low-quality resumes selected are then drawn at random to receive African-American names, the other high- and low-quality resumes receive White names.[24] We use male and female names for sales jobs, whereas we use nearly exclusively female names for administrative and clerical jobs to increase callback rates.[25] Based on sex, race, city, and resume quality, we assign a resume the appropriate phone number. We also select at random a postal address. Finally, e-mail addresses are added to most of the high-quality resumes.[26] The final resumes are formatted, with fonts, layout, and cover letter style chosen at random. The resumes are then faxed (or in a few cases mailed) to the employer. All in all, we respond to more than 1,300 employment ads over the entire sample period and send close to 5,000 resumes.

D. *Measuring Responses*

We measure whether a given resume elicits a callback or e-mail back for an interview. For each phone or e-mail response, we use the content of the message left by the employer (name of the applicant, company name, telephone number for contact) to match the response to the corresponding resume-ad pair.[27] Any attempt by employers to contact applicants via postal mail cannot be measured in our experiment since the addresses are fictitious. Several human resource managers confirmed to us that

[21] The e-mail addresses are registered on Yahoo.com, Angelfire.com, or Hotmail.com.

[22] This period spans tighter and slacker labor markets. In our data, this is apparent as callback rates (and number of new ads) dropped after September 11, 2001. Interestingly, however, the racial gap we measure is the same across these two periods.

[23] In some instances, our resume bank does not have four resumes that are appropriate matches for a given ad. In such instances, we send only two resumes.

[24] Though the same names are repeatedly used in our experiment, we guarantee that no given ad receives multiple resumes with the same name.

[25] Male names were used for a few administrative jobs in the first month of the experiment.

[26] In the first month of the experiment, a few high-quality resumes were sent without e-mail addresses and a few low-quality resumes were given e-mail addresses. See Table 3 for details.

[27] Very few employers used e-mail to contact an applicant back.

TABLE 1—MEAN CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES

	Percent callback for White names	Percent callback for African-American names	Ratio	Percent difference (*p*-value)
Sample:				
All sent resumes	9.65	6.45	1.50	3.20
	[2,435]	[2,435]		(0.0000)
Chicago	8.06	5.40	1.49	2.66
	[1,352]	[1,352]		(0.0057)
Boston	11.63	7.76	1.50	4.05
	[1,083]	[1,083]		(0.0023)
Females	9.89	6.63	1.49	3.26
	[1,860]	[1,886]		(0.0003)
Females in administrative jobs	10.46	6.55	1.60	3.91
	[1,358]	[1,359]		(0.0003)
Females in sales jobs	8.37	6.83	1.22	1.54
	[502]	[527]		(0.3523)
Males	8.87	5.83	1.52	3.04
	[575]	[549]		(0.0513)

Notes: The table reports, for the entire sample and different subsamples of sent resumes, the callback rates for applicants with a White-sounding name (column 1) an an African-American-sounding name (column 2), as well as the ratio (column 3) and difference (column 4) of these callback rates. In brackets in each cell is the number of resumes sent in that cell. Column 4 also reports the *p*-value for a test of proportion testing the null hypothesis that the callback rates are equal across racial groups.

employers rarely, if ever, contact applicants via postal mail to set up interviews.

E. *Weaknesses of the Experiment*

We have already highlighted the strengths of this experiment relative to previous audit studies. We now discuss its weaknesses. First, our outcome measure is crude, even relative to the previous audit studies. Ultimately, one cares about whether an applicant gets the job and about the wage offered conditional on getting the job. Our procedure, however, simply measures callbacks for interviews. To the extent that the search process has even moderate frictions, one would expect that reduced interview rates would translate into reduced job offers. However, we are not able to translate our results into gaps in hiring rates or gaps in earnings.

Another weakness is that the resumes do not directly report race but instead suggest race through personal names. This leads to various sources of concern. First, while the names are chosen to make race salient, some employers may simply not notice the names or not recognize their racial content. On a related note, because we are not assigning race but only race-specific names, our results are not representative of the average African-American (who may not have such a racially distinct name).[28] We return to this issue in Section IV, subsection B.

Finally, and this is an issue pervasive in both our study and the pair-matching audit studies, newspaper ads represent only one channel for job search. As is well known from previous work, social networks are another common means through which people find jobs and one that clearly cannot be studied here. This omission could qualitatively affect our results if African-Americans use social networks more or if employers who rely more on networks differentiate less by race.[29]

III. Results

A. *Is There a Racial Gap in Callback?*

Table 1 tabulates average callback rates by racial soundingness of names. Included in brackets under each rate is the number of resumes sent in that cell. Row 1 presents our results for the full data set. Resumes with White

[28] As Appendix Table A1 indicates, the African-American names we use are, however, quite common among African-Americans, making this less of a concern.

[29] In fact, there is some evidence that African-Americans may rely less on social networks for their job search (Harry J. Holzer, 1987).

names have a 9.65 percent chance of receiving a callback. Equivalent resumes with African-American names have a 6.45 percent chance of being called back. This represents a difference in callback rates of 3.20 percentage points, or 50 percent, that can solely be attributed to the name manipulation. Column 4 shows that this difference is statistically significant.[30] Put in other words, these results imply that a White applicant should expect on average one callback for every 10 ads she or he applies to; on the other hand, an African-American applicant would need to apply to about 15 different ads to achieve the same result.[31]

How large are these effects? While the cost of sending additional resumes might not be large per se, this 50-percent gap could be quite substantial when compared to the rate of arrival of new job openings. In our own study, the biggest constraining factor in sending more resumes was the limited number of new job openings each week. Another way to benchmark the measured return to a White name is to compare it to the returns to other resume characteristics. For example, in Table 5, we will show that, at the average number of years of experience in our sample, an extra year of experience increases the likelihood of a callback by a 0.4 percentage point. Based on this point estimate, the return to a White name is equivalent to about eight additional years of experience.

Rows 2 and 3 break down the full sample of sent resumes into the Boston and Chicago markets. About 20 percent more resumes were sent in Chicago than in Boston. The average callback rate (across races) is lower in Chicago than in Boston. This might reflect differences in labor market conditions across the two cities over the experimental period or maybe differences in the ability of the MIT and Chicago teams of research assistants in selecting resumes that were good matches for a given help-wanted ad. The percentage difference in callback rates is, however, strikingly similar across both cities. White applicants are 49 percent more likely than African-American applicants to receive a callback in Chicago and 50 percent more likely in Boston. These racial differences are statistically significant in both cities.

Finally, rows 4 to 7 break down the full sample into female and male applicants. Row 4 displays the average results for all female names while rows 5 and 6 break the female sample into administrative (row 5) and sales jobs (row 6); row 7 displays the average results for all male names. As noted earlier, female names were used in both sales and administrative job openings whereas male names were used close to exclusively for sales openings.[32] Looking across occupations, we find a significant racial gap in callbacks for both males (52 percent) and females (49 percent). Comparing males to females in sales occupations, we find a larger racial gap among males (52 percent versus 22 percent). Interestingly, females in sales jobs appear to receive more callbacks than males; however, this (reverse) gender gap is statistically insignificant and economically much smaller than any of the racial gaps discussed above.

Rather than studying the distribution of callbacks at the applicant level, one can also tabulate the distribution of callbacks at the employment-ad level. In Table 2, we compute the fraction of employers that treat White and African-American applicants equally, the fraction of employers that favor White applicants and the fraction of employers that favor African-American applicants. Because we send up to four resumes in response to each sampled ad, the three categories above can each take three different forms. Equal treatment occurs when either no applicant gets called back, one White and one African-American get called back or two Whites and two African-Americans get called back. Whites are favored when either only one White gets called back, two Whites and no African-American get called back or two Whites and one African-American get called back. African-Americans are favored in all other cases.

As Table 2 indicates, equal treatment occurs for about 88 percent of the help-wanted ads. As expected, the major source of equal treatment comes from the high fraction of ads for which

[30] These statistical tests assume independence of callbacks. We have, however, verified that the results stay significant when we assume that the callbacks are correlated either at the employer or first-name level.

[31] This obviously assumes that African-American applicants cannot assess a priori which firms are more likely to treat them more or less favorably.

[32] Only about 6 percent of all male resumes were sent in response to an administrative job opening.

TABLE 2—DISTRIBUTION OF CALLBACKS BY EMPLOYMENT AD

Equal Treatment:	No Callback	1W + 1B	2W + 2B
88.13 percent	83.37	3.48	1.28
[1,166]	[1,103]	[46]	[17]
Whites Favored (WF):	1W + 0B	2W + 0B	2W + 1B
8.39 percent	5.59	1.44	1.36
[111]	[74]	[19]	[18]
African-Americans Favored (BF):	1B + 0W	2B + 0W	2B + 1W
3.48 percent	2.49	0.45	0.53
[46]	[33]	[6]	[7]
Ho: *WF* = *BF*			
$p = 0.0000$			

Notes: This table documents the distribution of callbacks at the employment-ad level. "No Callback" is the percent of ads for which none of the fictitious applicants received a callback. "1W + 1B" is the percent of ads for which exactly one White and one African-American applicant received a callback. "2W + 2B" is the percent of ads for which exactly two White applicants and two African-American applicants received a callback. "Equal Treatment" is defined as the sum of "No Callback," "1W + 1B," and "2W + 2B." "1W + 0B" is the percent of ads for which exactly one White applicant and no African-American applicant received a call back. "2W + 0B" is the percent of ads for which excatly two White applicants and no African-American applicant received a callback. "2W + 1B" is the percent of ads for which exactly two White applicants and one African-American applicant received a callback. "Whites Favored" is defined as the sum of "1W + 0B," "2W + 0B," and "2W + 1B." "1B + 0W" is the percent of ads for which exactly one African-American applicant and no White applicant received a callback. "2B + 0W" is the percent of ads for which exactly two African-American applicants and no White applicant received a callback. "2B + 1W" is the percent of ads for which exactly two African-American applicants and one White applicant received a callback. "African-Americans Favored" is defined as the sum of "1B + 0W," "2B + 0W," and "2B + 1W." In brackets in each cell is the number of employment ads in that cell. "*Ho*: *WF* = *WB*" reports the *p*-value for a test of symmetry between the proportion of employers that favor White names and the proportion of employers that favor African-American names.

no callbacks are recorded (83 percent of the ads). Whites are favored by nearly 8.4 percent of the employers, with a majority of these employers contacting exactly one White applicant. African-Americans, on the other hand, are favored by only about 3.5 percent of employers. We formally test whether there is symmetry in the favoring of Whites over African-Americans and African-Americans over Whites. We find that the difference between the fraction of employers favoring Whites and the fraction of employers favoring African-Americans is statistically very significant ($p = 0.0000$).

B. *Do African-Americans Receive Different Returns to Resume Quality?*

Our results so far demonstrate a substantial gap in callback based on applicants' names. Next, we would like to learn more about the factors that may influence this gap. More specifically, we ask how employers respond to improvements in African-American applicants' credentials. To answer this question, we examine how the racial gap in callback varies by resume quality.

As we explained in Section II, for most of the employment ads we respond to, we send four different resumes: two higher-quality and two lower-quality ones. Table 3 gives a better sense of which factors enter into this subjective classification. Table 3 displays means and standard deviations of the most relevant resume characteristics for the full sample (column 1), as well as broken down by race (columns 2 and 3) and resume quality (columns 4 and 5). Since applicants' names are randomized, there is no difference in resume characteristics by race. Columns 4 and 5 document the objective differences between resumes subjectively classified as high and low quality. Higher-quality applicants have on average close to an extra year of labor market experience, fewer employment holes (where an employment hole is defined as a period of at least six months without a reported job), are more likely to have worked while at school, and to report some military experience. Also, higher-quality applicants are more likely to have an e-mail address, to have received some honors, and to list some computer skills and other special skills (such as a certification degree or foreign language skills) on their resume. Note that the higher- and lower-quality resumes do not differ on average with regard to

TABLE 3—RESUME CHARACTERISTICS: SUMMARY STATISTICS

Sample:	All resumes	White names	African-American	Higher quality	Lower quality
Characteristic:					
College degree	0.72	0.72	0.72	0.72	0.71
(Y = 1)	(0.45)	(0.45)	(0.45)	(0.45)	(0.45)
Years of experience	7.84	7.86	7.83	8.29	7.39
	(5.04)	(5.07)	(5.01)	(5.29)	(4.75)
Volunteering experience?	0.41	0.41	0.41	0.79	0.03
(Y = 1)	(0.49)	(0.49)	(0.49)	(0.41)	(0.16)
Military experience?	0.10	0.09	0.10	0.19	0.00
(Y = 1)	(0.30)	(0.29)	(0.30)	(0.39)	(0.06)
E-mail address?	0.48	0.48	0.48	0.92	0.03
(Y = 1)	(0.50)	(0.50)	(0.50)	(0.27)	(0.17)
Employment holes?	0.45	0.45	0.45	0.34	0.56
(Y = 1)	(0.50)	(0.50)	(0.50)	(0.47)	(0.50)
Work in school?	0.56	0.56	0.56	0.72	0.40
(Y = 1)	(0.50)	(0.50)	(0.50)	(0.45)	(0.49)
Honors?	0.05	0.05	0.05	0.07	0.03
(Y = 1)	(0.22)	(0.23)	(0.22)	(0.25)	(0.18)
Computer skills?	0.82	0.81	0.83	0.91	0.73
(Y = 1)	(0.38)	(0.39)	(0.37)	(0.29)	(0.44)
Special skills?	0.33	0.33	0.33	0.36	0.30
(Y = 1)	(0.47)	(0.47)	(0.47)	(0.48)	(0.46)
Fraction high school dropouts in applicant's zip code	0.19	0.19	0.19	0.19	0.18
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Fraction college or more in applicant's zip code	0.21	0.21	0.21	0.21	0.22
	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
Fraction Whites in applicant's zip code	0.54	0.55	0.54	0.53	0.55
	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)
Fraction African-Americans in applicant's zip code	0.31	0.31	0.31	0.32	0.31
	(0.33)	(0.33)	(0.33)	(0.33)	(0.33)
Log(median per capital income) in applicant's zip code	9.55	9.55	9.55	9.54	9.56
	(0.56)	(0.56)	(0.55)	(0.54)	(0.57)
Sample size	4,870	2,435	2,435	2,446	2,424

Notes: The table reports means and standard deviations for the resume characteristics as listed on the left. Column 1 refers to all resumes sent; column 2 refers to resumes with White names; column 3 refers to resumes with African-American names; column 4 refers to higher-quality resumes; column 5 refers to lower-quality resumes. See text for details.

applicants' education level. This reflects the fact that all sent resumes, whether high or low quality, are chosen to be good matches for a given job opening. About 70 percent of the sent resumes report a college degree.[33]

The last five rows of Table 3 show summary characteristics of the applicants' zip code address. Using 1990 Census data, we compute the fraction of high school dropouts, fraction of college educated or more, fraction of Whites, fraction of African-Americans and log(median per capital income) for each zip code used in the experiment. Since addresses are randomized within cities, these neighborhood quality measures are uncorrelated with race or resume quality.

The differences in callback rates between high- and low-quality resumes are presented in Panel A of Table 4. The first thing to note is that the resume quality manipulation works: higher-quality resumes receive more callbacks. As row 1 indicates, we record a callback rate of close to 11 percent for White applicants with a higher-quality resume, compared to 8.5 percent for White applicants with lower-quality resumes. This is a statistically significant difference of 2.29 percentage points, or 27 percent ($p = 0.0557$). Most strikingly, African-Americans experience much less of an increase in callback

[33] This varies from about 50 percent for the clerical and administrative support positions to more than 80 percent for the executive, managerial, and sales representatives positions.

TABLE 4—AVERAGE CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES AND RESUME QUALITY

	Panel A: Subjective Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (p-value)
White names	8.50	10.79	1.27	2.29
	[1,212]	[1,223]		(0.0557)
African-American names	6.19	6.70	1.08	0.51
	[1,212]	[1,223]		(0.6084)
	Panel B: Predicted Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (p- value)
White names	7.18	13.60	1.89	6.42
	[822]	[816]		(0.0000)
African-American names	5.37	8.60	1.60	3.23
	[819]	[814]		(0.0104)

Notes: Panel A reports the mean callback percents for applicant with a White name (row 1) and African-American name (row 2) depending on whether the resume was subjectively qualified as a lower quality or higher quality. In brackets is the number of resumes sent for each race/quality group. The last column reports the p-value of a test of proportion testing the null hypothesis that the callback rates are equal across quality groups within each racial group. For Panel B, we use a third of the sample to estimate a probit regression of the callback dummy on the set of resume characteristics as displayed in Table 3. We further control for a sex dummy, a city dummy, six occupation dummies, and a vector of dummy variables for job requirements as listed in the employment ad (see Section III, subsection D, for details). We then use the estimated coefficients on the set of resume characteristics to estimate a predicted callback for the remaining resumes (two-thirds of the sample). We call "high-quality" resumes the resumes that rank above the median predicted callback and "low-quality" resumes the resumes that rank below the median predicted callback. In brackets is the number of resumes sent for each race/quality group. The last column reports the p-value of a test of proportion testing the null hypothesis that the callback percents are equal across quality groups within each racial group.

rate for similar improvements in their credentials. African-Americans with higher-quality resumes receive a callback 6.7 percent of the time, compared to 6.2 percent for African-Americans with lower quality resumes. This is only a 0.51-percentage-point, or 8-percent, difference and this difference is not statistically significant ($p = 0.6084$).

Instead of relying on the subjective quality classification, Panel B directly uses resume characteristics to classify the resumes. More specifically, we use a random subsample of one-third of the resumes to estimate a probit regression of the callback dummy on the resume characteristics listed in Table 3. We further control for a sex dummy, a city dummy, six occupation dummies, and a vector of job requirements as listed in the employment ads.[34] We then use the estimated coefficients on the resume characteristics to rank the remaining two-thirds of the resumes by predicted callback. In Panel B, we classify as "high" those resumes that have above-median-predicted callback; similarly, we classify as "low" those resumes that have below-median-predicted callback. As one can see from Panel B, qualitatively similar results emerge from this analysis. While African-Americans do appear to significantly benefit from higher-quality resumes under this alternative classification, they benefit less than Whites. The ratio of callback rates for high- versus low-quality resumes is 1.60 for African Americans, compared to 1.89 for Whites.

In Table 5, we directly report the results of race-specific probit regressions of the callback dummy on resume characteristics. We, however, start in column 1 with results for the full sample of sent resumes. As one can see, many of the resume characteristics have the expected effect on the likelihood of a callback. The addition of an e-mail address, honors, and special skills all have a positive and significant effect on the likelihood of a callback.[35] Also, more experienced applicants are more likely to get called back: at the average number of years of experience in our sample (eight years), each

[34] See Section III, subsection D, for more details on these occupation categories and job requirements.

[35] Note that the e-mail address dummy, because it is close to perfectly correlated with the subjective resume-quality variable, may in part capture some other unmeasured resume characteristics that may have led us to categorize a given resume as higher quality.

TABLE 5—EFFECT OF RESUME CHARACTERISTICS ON LIKELIHOOD OF CALLBACK

Dependent Variable: Callback Dummy Sample:	All resumes	White names	African-American names
Years of experience (*10)	0.07	0.13	0.02
	(0.03)	(0.04)	(0.03)
Years of experience^2 (*100)	−0.02	−0.04	−0.00
	(0.01)	(0.01)	(0.01)
Volunteering? (Y = 1)	−0.01	−0.01	0.01
	(0.01)	(0.01)	(0.01)
Military experience? (Y = 1)	−0.00	0.02	−0.01
	(0.01)	(0.03)	(0.02)
E-mail? (Y = 1)	0.02	0.03	−0.00
	(0.01)	(0.01)	(0.01)
Employment holes? (Y = 1)	0.02	0.03	0.01
	(0.01)	(0.02)	(0.01)
Work in school? (Y = 1)	0.01	0.02	−0.00
	(0.01)	(0.01)	(0.01)
Honors? (Y = 1)	0.05	0.06	0.03
	(0.02)	(0.03)	(0.02)
Computer skills? (Y = 1)	−0.02	−0.04	−0.00
	(0.01)	(0.02)	(0.01)
Special skills? (Y = 1)	0.05	0.06	0.04
	(0.01)	(0.02)	(0.01)
Ho: Resume characteristics effects are all zero (*p*-value)	54.50 (0.0000)	57.59 (0.0000)	23.85 (0.0080)
Standard deviation of predicted callback	0.047	0.062	0.037
Sample size	4,870	2,435	2,435

Notes: Each column gives the results of a probit regression where the dependent variable is the callback dummy. Reported in the table are estimated marginal changes in probability for the continuous variables and estimated discrete changes for the dummy variables. Also included in each regression are a city dummy, a sex dummy, six occupation dummies, and a vector of dummy variables for job requirements as listed in the employment ad (see Section III, subsection D, for details). Sample in column 1 is the entire set of sent resumes; sample in column 2 is the set of resumes with White names; sample in column 3 is the set of resumes with African-American names. Standard errors are corrected for clustering of the observations at the employment-ad level. Reported in the second to last row are the *p*-values for a χ^2 testing that the effects on the resume characteristics are all zero. Reported in the second to last row is the standard deviation of the predicted callback rate.

extra year of experience increases the likelihood of a callback by about a 0.4 percentage point. The most counterintuitive effects come from computer skills, which appear to negatively predict callback, and employment holes, which appear to positively predict callback.

The same qualitative patterns hold in column 2 where we focus on White applicants. More importantly, the estimated returns to an e-mail address, additional work experience, honors, and special skills appear economically stronger for that racial group. For example, at the average number of years of experience in our sample, each extra year of experience increases the likelihood of a callback by about a 0.7 percentage point.

As might have been expected from the two previous columns, we find that the estimated returns on these resume characteristics are all economically and statistically weaker for African-American applicants (column 3). In fact, all the estimated effects for African-Americans are statistically insignificant, except for the return to special skills. Resume characteristics thus appear less predictive of callback rates for African-Americans than they are for Whites. To illustrate this more saliently, we predict callback rates using either regression estimates in column 2 or regression estimates in column 3. The standard deviation of the predicted callback from column 2 is 0.062, whereas it is only 0.037 from column 3. In summary, employers simply seem to pay less attention or discount more the characteristics listed on the

TABLE 6—EFFECT OF APPLICANT'S ADDRESS ON LIKELIHOOD OF CALLBACK

Dependent Variable: Callback Dummy						
Zip code characteristic:	Fraction Whites		Fraction college or more		Log(per capital income)	
Zip code characteristic	0.020	0.020	0.054	0.053	0.018	0.014
	(0.012)	(0.016)	(0.022)	(0.031)	(0.007)	(0.010)
Zip code characteristic*	—	−0.000	—	−0.002	—	0.008
African-American name		(0.024)		(0.048)		(0.015)
African-American name	—	−0.031	—	−0.031	—	−0.112
		(0.015)		(0.013)		(0.152)

Notes: Each column gives the results of a probit regression where the dependent variable is the callback dummy. Reported in the table is the estimated marginal change in probability. Also included in columns 1, 3, and 5 is a city dummy; also included in columns 2, 4, and 6 is a city dummy and a city dummy interacted with a race dummy. Standard errors are corrected for clustering of the observations at the employment-ad level.

resumes with African-American-sounding names. Taken at face value, these results suggest that African-Americans may face relatively lower individual incentives to invest in higher skills.[36]

C. *Applicants' Address*

An incidental feature of our experimental design is the random assignment of addresses to the resumes. This allows us to examine whether and how an applicant's residential address, all else equal, affects the likelihood of a callback. In addition, and most importantly for our purpose, we can also ask whether African-American applicants are helped relatively more by residing in more affluent neighborhoods.

We perform this analysis in Table 6. We start (columns 1, 3, and 5) by discussing the effect of neighborhood of residence across all applicants. Each of these columns reports the results of a probit regression of the callback dummy on a specific zip code characteristic and a city dummy. Standard errors are corrected for clustering of the observations at the employment-ad level. We find a positive and significant effect of neighborhood quality on the likelihood of a callback. Applicants living in Whiter (column 1), more educated (column 3), or higher-income (column 5) neighborhoods have a higher probability of receiving a callback. For example, a 10-percentage-point increase in the fraction of college-educated in zip code of residence increases the likelihood of a callback by a 0.54 percentage point (column 3).

In columns 2, 4, and 6, we further interact the zip code characteristic with a dummy variable for whether the applicant is African-American or not. Each of the probit regressions in these columns also includes an African-American dummy, a city dummy, and an interaction of the city dummy with the African-American dummy. There is no evidence that African-Americans benefit any more than Whites from living in a Whiter, more educated zip code. The estimated interactions between fraction White and fraction college educated with the African-American dummy are economically very small and statistically insignificant. We do find an economically more meaningful effect of zip code median income level on the racial gap in callback; this effect, however, is statistically insignificant.

In summary, while neighborhood quality affects callbacks, African-Americans do not benefit more than Whites from living in better neighborhoods. If ghettos and bad neighborhoods are particularly stigmatizing for African-Americans, one might have expected African-Americans to be helped more by having a "better" address. Our results do not support this hypothesis.

D. *Job and Employer Characteristics*

Table 7 studies how various job requirements (as listed in the employment ads) and employer characteristics correlate with the racial gap in callback. Each row of Table 7 focuses on a specific job or employer characteristic, with

[36] This of course assumes that the changes in job and wage offers associated with higher skills are the same across races, or at least not systematically larger for African-Americans.

TABLE 7—EFFECT OF JOB REQUIREMENT AND EMPLOYER CHARACTERISTICS ON RACIAL DIFFERENCES IN CALLBACKS

Job requirement:	Sample mean (standard deviation)	Marginal effect on callbacks for African-American names
Any requirement? (Y = 1)	0.79	0.023
	(0.41)	(0.015)
Experience? (Y = 1)	0.44	0.011
	(0.49)	(0.013)
Computer skills? (Y = 1)	0.44	0.000
	(0.50)	(0.013)
Communication skills? (Y = 1)	0.12	−0.000
	(0.33)	(0.015)
Organization skills? (Y = 1)	0.07	0.028
	(0.26)	(0.029)
Education? (Y = 1)	0.11	−0.031
	(0.31)	(0.017)
Total number of requirements	1.18	0.002
	(0.93)	(0.006)
Employer characteristic:	**Sample mean (standard deviation)**	**Marginal effect on callbacks for African-American names**
Equal opportunity employer? (Y = 1)	0.29	−0.013
	(0.45)	(0.012)
Federal contractor? (Y = 1)	0.11	−0.035
(N = 3,102)	(0.32)	(0.016)
Log(employment)	5.74	−0.001
(N = 1,690)	(1.74)	(0.005)
Ownership status:		
(N = 2,878)		
Privately held	0.74	0.011
		(0.019)
Publicly traded	0.15	−0.025
		(0.015)
Not-for-profit	0.11	0.025
		(0.042)
Fraction African-Americans in employer's zip code	0.08	0.117
(N = 1,918)	(0.15)	(0.062)

Notes: Sample is all sent resumes (N = 4,870) unless otherwise specified in column 1. Column 2 reports means and standard deviations (in parentheses) for the job requirement or employer characteristic. For ads listing an experience requirement, 50.1 percent listed "some," 24.0 percent listed "two years or less," and 25.9 percent listed "three years or more." For ads listing an education requirement, 8.8 percent listed a high school degree, 48.5 percent listed some college, and 42.7 percent listed at least a four-year college degree. Column 3 reports the marginal effect of the job requirement or employer characteristic listed in that row on differential treatment. Specifically, each cell in column 3 corresponds to a different probit regression of the callback dummy on an African-American name dummy, a dummy for the requirement or characteristic listed in that row and the interaction of the requirement or characteristic dummy with the African-American name dummy. Reported in each cell is the estimated change in probability for the interaction term. Standard errors are corrected for clustering of the observations at the employment-ad level.

summary statistics in column 2. Column 3 shows the results of various probit regressions. Each entry in this column is the marginal effect of the specific characteristic listed in that row on the racial gap in callback. More specifically, each entry is from a separate probit regression of a callback dummy on an African-American dummy, the characteristic listed in that row and the interaction of that characteristic with the African-American dummy. The reported coefficient is that on the interaction term.

We start with job requirements. About 80 percent of the ads state some form of requirement. About 44 percent of the ads require some minimum experience, of which roughly 50 percent simply ask for "some experience," 24 percent less than two years, and 26 percent at least three years of experience. About 44 percent of

ads mention some computer knowledge requirement, which can range from Excel or Word to more esoteric software programs. Good communication skills are explicitly required in about 12 percent of the ads. Organization skills are mentioned 7 percent of the time. Finally, only about 11 percent of the ads list an explicit education requirement. Of these, 8.8 percent require a high school degree, 48.5 percent some college (such as an associate degree), and the rest at least a four-year college degree.[37]

Despite this variability, we find little systematic relationship between any of the requirements and the racial gap in callback. The point estimates in column 3 show no consistent economic pattern and are all statistically weak. Measures of job quality, such as experience or computer skills requirements, do not predict the extent of the racial gap. Communication or other interpersonal skill requirements have no effect on the racial gap either.[38]

We also study employer characteristics. Collecting such information is a more difficult task since it is not readily available from the employment ads we respond to. The only piece of employer information we can directly collect from the employment ad is whether or not the employer explicitly states being an "Equal Opportunity Employer." In several cases, the name of the employer is not even mentioned in the ad and the only piece of information we can rely on is the fax number which applications must be submitted to. We therefore have to turn to supplemental data sources. For employment ads that do not list a specific employer, we first use the fax number to try to identify the company name via Web reverse-lookup services. Based on company names, we use three different data sources (*Onesource Business Browser, Thomas Register,* and *Dun and Bradstreet Million Dollar Directory, 2001*) to track company information such as total employment, industry, and ownership status. Using this same set of data sources, we also try to identify the specific zip code of the company (or company branch) that resumes are to be sent to. Finally, we use the Federal Procurement and Data Center Web site to find a list of companies that have federal contracts.[39] The racial difference in callback rates for the subsamples where employer characteristics could be determined is very similar in magnitude to that in the full sample.

Employer characteristics differ significantly across ads. Twenty-nine percent of all employers explicitly state that they are "Equal Opportunity Employers." Eleven percent are federal contractors and, therefore, might face greater scrutiny under affirmative action laws. The average company size is around 2,000 employees but there is a lot of variation across firms. Finally, 74 percent of the firms are privately held, 15 percent are publicly traded, and 11 percent are not-for-profit organizations.

Neither "Equal Opportunity Employers" nor federal contractors appear to treat African-Americans more favorably. In fact, each of these employer characteristics is associated with a larger racial gap in callback (and this effect is marginally significant for federal contractors). Differential treatment does not vary with employer size.[40] Point estimates indicate less differential treatment in the not-for-profit sector; however, this effect is very noisily estimated.[41]

In an unpublished Appendix (available from the authors upon request), we also study how the racial gap in callback varies by occupation and industry. Based on the employment ad listings, we classify the job openings into six occupation categories: executives and managers; administrative supervisors; sales representatives; sales workers; secretaries and legal assistants; clerical workers. We also, when possible,

[37] Other requirements sometimes mentioned include typing skills for secretaries (with specific words-per-minute minimum thresholds), and, more rarely, foreign language skills.

[38] Other ways of estimating these effects produce a similar nonresult. Among other things, we considered including a city dummy or estimating the effects separately by city; we also estimated one single probit regression including all requirements at once.

[39] This Web site (www.fpdc.gov) is accurate up to and including March 21, 2000.

[40] Similar results hold when we measure employer size using a total sales measure rather than an employment measure.

[41] Our measurement of the racial gap by firm or employer type may not be a good indicator of the fraction of African-Americans actually employed in these firms. For example, "Equal Opportunity Employers" may receive a higher fraction of African-American resumes. Their actual hiring may therefore look different from that of non "Equal Opportunity Employers" when one considers the full set of resumes they receive.

classify employers into six industry categories: manufacturing; transportation and communication; wholesale and retail trade; finance, insurance, and real estate; business and personal services; health, educational, and social services. We then compute occupation and industry-specific racial gaps in callback and relate these gaps to 1990 Census-based measures of occupation and industry earnings, as well as Census-based measures of the White/African-American wage gap in these occupations and industries.

We find a positive White/African-American gap in callbacks in all occupation and industry categories (except for transportation and communication). While average earnings vary a lot across the occupations covered in the experiment, we find no systematic relationship between occupation earnings and the racial gap in callback. Similarly, the industry-specific gaps in callback do not relate well to a measure of inter-industry wage differentials. In fact, while the racial gap in callback rates varies somewhat across occupations and industries, we cannot reject the null hypothesis that the gap is the same across all these categories.

The last row of Table 7 focuses on the marginal effect of employer location on the racial gap in callback.[42] We use as a measure of employer location the zip code of the company (or company branch) resumes were to be sent to. More specifically, we ask whether differential treatment by race varies with the fraction of African-Americans in the employer's zip code. We find a marginally significant positive effect of employer location on African-American callbacks but this effect is extremely small. In regressions not reported here (but available from the authors upon request), we reestimate this effect separately by city. While the point estimates are positive for both cities, the effect is only statistically significant for Chicago.

IV. Interpretation

Three main sets of questions arise when interpreting the results above. First, does a higher callback rate for White applicants imply that employers are discriminating against African-Americans? Second, does our design only isolate the effect of race or is the name manipulation conveying some other factors than race? Third, how do our results relate to different models of racial discrimination?

A. *Interpreting Callback Rates*

Our results indicate that for two identical individuals engaging in an identical job search, the one with an African-American name would receive fewer interviews. Does differential treatment within our experiment imply that employers are discriminating against African-Americans (whether it is rational, prejudice-based, or other form of discrimination)? In other words, could the lower callback rate we record for African-American resumes *within our experiment* be consistent with a racially neutral review of the *entire pool* of resumes the surveyed employers receive?

In a racially neutral review process, employers would rank order resumes based on their quality and call back all applicants that are above a certain threshold. Because names are randomized, the White and African-American resumes we send should rank similarly on average. So, irrespective of the skill and racial composition of the applicant pool, a race-blind selection rule would generate equal treatment of Whites and African-Americans. So our results must imply that employers use race as a factor when reviewing resumes, which matches the legal definition of discrimination.

But even rules where employers are not trying to interview as few African-American applicants as possible may generate observed differential treatment in our experiment. One such hiring rule would be employers trying to interview a target level of African-American candidates. For example, perhaps the average firm in our experiment aims to produce an interview pool that matches the population base rate. This rule could produce the observed differential treatment if the average firm receives a higher proportion of African-American resumes than the population base rate because African-Americans disproportionately apply to the jobs and industries in our sample.[43]

[42] For previous work on the effect of employer location on labor market discrimination, see, for example, Steven Raphael et al. (2000).

[43] Another variant of this argument is that the (up to) two African-American resumes we sent are enough to signifi-

Some of our other findings may be consistent with such a rule. For example, the fact that "Equal Opportunity Employers" or federal contractors do not appear to discriminate any less may reflect the fact that such employers receive more applications from African-Americans. On the other hand, other key findings run counter to this rule. As we discuss above, we find no systematic difference in the racial gap in callback across occupational or industry categories, despite the large variation in the fraction of African-Americans looking for work in those categories. African-Americans are underrepresented in managerial occupations, for example. If employers matched base rates in the population, the few African-Americans who apply to these jobs should receive a higher callback rate than Whites. Yet, we find that the racial gap in managerial occupations is the same as in all the other job categories. This rule also runs counter to our findings on returns to skill. Suppose firms are struggling to find White applicants but overwhelmed with African-American ones. Then they should be less sensitive to the quality of White applicants (as they are trying to fill in their hiring quota for Whites) and much more sensitive to the quality of Black applicants (when they have so many to pick from). Thus, it is unlikely that the differential treatment we observe is generated by hiring rules such as these.

cantly distort the racial composition of the entire applicant pool. This is unlikely for two reasons. First, anecdotal evidence and the empirically low callback rates we record suggest that firms typically receive many hundreds of resumes in response to each ad they post. Hence, the (up to) four resumes we send out are unlikely to influence the racial composition of the pool. Second, the similar racial gap in callback we observe across the two cities goes counter to this interpretation since the racial composition base rates differ quite a lot across these two cities. Another variant of this argument is that, for some reason, the average firm in our sample receives a lot of high-quality resumes from African-American applicants and much fewer high-quality resumes from White applicants. Hypothetically, this might occur if high-quality African-Americans are much more likely to use help-wanted ads rather than other job search channels. If employers perform within-race comparisons and again want to target a certain racial mix in their interviewing and hiring, our African-American resumes may naturally receive lower callbacks as they are competing with many more high-quality applicants. This specific argument would be especially relevant in a case where the average sampled employer is "known" to be good to African-Americans. But our selection procedure for the employment ads did not allow for such screening: we simply responded to as many ads as possible in the targeted occupational categories.

B. *Potential Confounds*

While the names we have used in this experiment strongly signal racial origin, they may also signal some other personal trait. More specifically, one might be concerned that employers are inferring social background from the personal name. When employers read a name like "Tyrone" or "Latoya," they may assume that the person comes from a disadvantaged background.[44] In the extreme form of this social background interpretation, employers do not care at all about race but are discriminating only against the social background conveyed by the names we have chosen.[45]

While plausible, we feel that some of our earlier results are hard to reconcile with this interpretation. For example, in Table 6, we found that while employers value "better" addresses, African-Americans are not helped more than Whites by living in Whiter or more educated neighborhoods. If the African-American names we have chosen mainly signal negative social background, one might have expected the estimated name gap to be lower for better addresses. Also, if the names mainly signal social background, one might have expected the name gap to be higher for jobs that rely more on soft skills or require more interpersonal interactions. We found no such evidence in Table 7.

We, however, directly address this alternative interpretation by examining the average social background of babies born with the names used in the experiment. We were able to obtain birth certificate data on mother's education (less than high school, high school or more) for babies born in Massachusetts between 1970 and

[44] Roland Fryer and Steven Levitt (2003) provide a recent analysis of social background and naming conventions amongst African-Americans.

[45] African-Americans as a whole come from more disadvantaged backgrounds than Whites. For this social class effect to be something of independent interest, one must assert that African-Americans with the African-American names we have selected are from a lower social background than the average African-American and/or that Whites with the White names we have selected are from a higher social background than the average White. We come back to this point below.

TABLE 8—CALLBACK RATE AND MOTHER'S EDUCATION BY FIRST NAME

White female			African-American female		
Name	Percent callback	Mother education	Name	Percent callback	Mother education
Emily	7.9	96.6	Aisha	2.2	77.2
Anne	8.3	93.1	Keisha	3.8	68.8
Jill	8.4	92.3	Tamika	5.5	61.5
Allison	9.5	95.7	Lakisha	5.5	55.6
Laurie	9.7	93.4	Tanisha	5.8	64.0
Sarah	9.8	97.9	Latoya	8.4	55.5
Meredith	10.2	81.8	Kenya	8.7	70.2
Carrie	13.1	80.7	Latonya	9.1	31.3
Kristen	13.1	93.4	Ebony	9.6	65.6
Average		91.7	Average		61.0
Overall		83.9	Overall		70.2
Correlation	−0.318	(p = 0.404)	Correlation	−0.383	(p = 0.309)
White male			African-American male		
Name	Percent callback	Mother education	Name	Percent callback	Mother education
Todd	5.9	87.7	Rasheed	3.0	77.3
Neil	6.6	85.7	Tremayne	4.3	—
Geoffrey	6.8	96.0	Kareem	4.7	67.4
Brett	6.8	93.9	Darnell	4.8	66.1
Brendan	7.7	96.7	Tyrone	5.3	64.0
Greg	7.8	88.3	Hakim	5.5	73.7
Matthew	9.0	93.1	Jamal	6.6	73.9
Jay	13.4	85.4	Leroy	9.4	53.3
Brad	15.9	90.5	Jermaine	9.6	57.5
Average		91.7	Average		66.7
Overall		83.5	Overall		68.9
Correlation	−0.0251	(p = 0.949)	Correlation	−0.595	(p = 0.120)

Notes: This table reports, for each first name used in the experiment, callback rate and average mother education. Mother education for a given first name is defined as the percent of babies born with that name in Massachusetts between 1970 and 1986 whose mother had at least completed a high school degree (see text for details). Within each sex/race group, first names are ranked by increasing callback rate. "Average" reports, within each race-gender group, the average mother education for all the babies born with one of the names used in the experiment. "Overall" reports, within each race-gender group, average mother education for all babies born in Massachusetts between 1970 and 1986 in that race-gender group. "Correlation" reports the Spearman rank order correlation between callback rate and mother education *within* each race-gender group as well as the p-value for the test of independence.

1986.[46] For each first name in our experiment, we compute the fraction of babies with that name and, in that gender-race cell, whose mothers have at least completed a high school degree.

In Table 8, we display the average callback rate for each first name along with this proxy for social background. Within each race-gender group, the names are ranked by increasing callback rate. Interestingly, there is significant

[46] This longer time span (compared to that used to assess name frequencies) was imposed on us for confidentiality reasons. When fewer than 10 births with education data available are recorded in a particular education-name cell, the exact number of births in that cell is not reported and we impute five births. Our results are not sensitive to this imputation. One African-American female name (Latonya) and two male names (Rasheed and Hakim) were imputed in this way. One African-American male name (Tremayne) had too few births with available education data and was therefore dropped from this analysis. Our results are qualitatively similar when we use a larger data set of California births for the years 1989 to 2000 (kindly provided to us by Steven Levitt).

variation in callback rates by name. Of course, chance alone could produce such variation because of the rather small number of observations in each cell (about 200 for the female names and 70 for the male names).[47]

The row labeled "Average" reports the average fraction of mothers that have at least completed high school for the set of names listed in that gender-race group. The row labeled "Overall" reports the average fraction of mothers that have at least completed high school for the full sample of births in that gender-race group. For example, 83.9 percent of White female babies born between 1970 and 1986 have mothers with at least a high school degree; 91.7 percent of the White female babies with one of the names used in the experiment have mothers with at least a high school degree.

Consistent with a social background interpretation, the African-American names we have chosen fall below the African-American average. For African-American male names, however, the gap between the experimental names and the population average is negligible. For White names, both the male and female names are above the population average.

But, more interestingly to us, there is substantial between-name heterogeneity in social background. African-American babies named Kenya or Jamal are affiliated with much higher mothers' education than African-American babies named Latonya or Leroy. Conversely, White babies named Carrie or Neil have lower social background than those named Emily or Geoffrey. This allows for a direct test of the social background hypothesis within our sample: are names associated with a worse social background discriminated against more? In the last row in each gender-race group, we report the rank-order correlation between callback rates and mother's education. The social background hypothesis predicts a positive correlation. Yet, for all four categories, we find the exact opposite. The *p*-values indicate that we cannot reject independence at standard significance levels except in the case of African-American males where we can almost reject it at the 10-percent level ($p = 0.120$). In summary, this test suggests little evidence that social background drives the measured race gap.

Names might also influence our results through familiarity. One could argue that the African-American names used in the experiment simply appear odd to human resource managers and that any odd name is discriminated against. But as noted earlier, the names we have selected are not particularly uncommon among African-Americans (see Appendix Table A1). We have also performed a similar exercise to that of Table 8 and measured the rank-order correlation between name-specific callback rates and name frequency within each gender-race group. We found no systematic positive correlation.

There is one final potential confound to our results. Perhaps what appears as a bias against African-Americans is actually the result of *reverse discrimination.* If qualified African-Americans are thought to be in high demand, then employers with average quality jobs might feel that an equally talented African-American would never accept an offer from them and thereby never call her or him in for an interview. Such an argument might also explain why African-Americans do not receive as strong a return as Whites to better resumes, since higher qualification only strengthens this argument. But this interpretation would suggest that among the better jobs, we ought to see evidence of reverse discrimination, or at least a smaller racial gap. However, as we discussed in Section III, subsection D, we do not find any such evidence. The racial gap does not vary across jobs with different skill requirements, nor does it vary across occupation categories. Even among the better jobs in our sample, we find that employers significantly favor applicants with White names.[48]

[47] We formally tested whether this variation was significant by estimating a probit regression of the callback dummy on all the personal first names, allowing for clustering of the observations at the employment-ad level. For all but African-American females, we cannot reject the null hypothesis that all the first name effects in the same race-gender group are the same. Of course, a lack of a rejection does not mean there is no underlying pattern in the between-name variation in callbacks that might have been detectable with larger sample sizes.

[48] One might argue that employers who reverse-discriminate hire through less formal channels than help-wanted ads. But this would imply that African-Americans are less likely to find jobs through formal channels. The evidence on exit out of unemployment does not paint a clear picture in this direction (Holzer, 1987).

C. *Relation to Existing Theories*

What do these results imply for existing models of discrimination? Economic theories of discrimination can be classified into two main categories: taste-based and statistical discrimination models.[49] Both sets of models can obviously "explain" our average racial gap in callbacks. But can these models explain our other findings? More specifically, we discuss the relevance of these models with a focus on two of the facts that have been uncovered in this paper: (i) the lower returns to credentials for African-Americans; (ii) the relative uniformity of the race gap across occupations, job requirements and, to a lesser extent, employer characteristics and industries.

Taste-based models (Gary S. Becker, 1961) differ in whose prejudiced "tastes" they emphasize: customers, coworkers, or employers. Customer and co-worker discrimination models seem at odds with the lack of significant variation of the racial gap by occupation and industry categories, as the amount of customer contact and the fraction of White employees vary quite a lot across these categories. We do not find a larger racial gap among jobs that explicitly require "communication skills" and jobs for which we expect either customer or coworker contacts to be higher (retail sales for example).

Because we do not know what drives employer tastes, employer discrimination models could be consistent with the lack of occupation and industry variation. Employer discrimination also matches the finding that employers located in more African-American neighborhoods appear to discriminate somewhat less. However, employer discrimination models would struggle to explain why African-Americans get relatively lower returns to their credentials. Indeed, the cost of indulging the discrimination taste should increase as the minority applicants' credentials increase.[50]

Statistical discrimination models are the prominent alternative to the taste-based models in the economics literature. In one class of statistical discrimination models, employers use (observable) race to proxy for *unobservable* skills (e.g., Edmund S. Phelps, 1972; Kenneth J. Arrow, 1973). This class of models struggle to explain the credentials effect as well. Indeed, the added credentials should lead to a larger update for African-Americans and hence greater returns to skills for that group.

A second class of statistical discrimination models "emphasize the precision of the information that employers have about individual productivity" (Altonji and Blank, 1999). Specifically, in these models, employers believe that the same observable signal is more precise for Whites than for African-Americans (Dennis J. Aigner and Glenn G. Cain, 1977; Shelly J. Lundberg and Richard Startz, 1983; Bradford Cornell and Ivo Welch, 1996). Under such models, African-Americans receive lower returns to observable skills because employers place less weight on these skills. However, how reasonable is this interpretation for our experiment? First, it is important to note that we are using the same set of resume characteristics for both racial groups. So the lower precision of information for African-Americans cannot be that, for example, an employer does not know what a high school degree from a very African-American neighborhood means (as in Aigner and Cain, 1977). Second, many of the credentials on the resumes are in fact externally and easily verifiable, such as a certification for a specific software.

An alternative version of these models would rely on bias in the observable signal rather than differential variance or noise of these signals by race. Perhaps the skills of African-Americans are discounted because affirmative action makes it easier for African-Americans to get these skills. While this is plausible for credentials such as an employee-of-the-month honor, it is unclear why this would apply to more verifiable and harder skills. It is equally unclear why work experience would be less rewarded since our study suggests that getting a job is more, not less, difficult for African-Americans.

The uniformity of the racial gap across occupations is also troubling for a statistical discrimination interpretation. Numerous factors that should affect the level of statistical discrimination, such as the importance of unobservable skills, the observability of qualifications, the precision of observable skills and the ease of

[49] Darity and Mason (1998) provide a more thorough review of a variety of economic theories of discrimination.

[50] One could, however, assume that employer tastes differ not just by race but also by race and skill, so that employers have greater prejudice against minority workers with better credentials. But the opposite preferences, employers having a particular distaste for low-skilled African-Americans, also seem reasonable.

performance measurement, may vary quite a lot across occupations.

This discussion suggests that perhaps other models may do a better job at explaining our findings. One simple alternative model is lexicographic search by employers. Employers receive so many resumes that they may use quick heuristics in reading these resumes. One such heuristic could be to simply read no further when they see an African-American name. Thus they may never see the skills of African-American candidates and this could explain why these skills are not rewarded. This might also to some extent explain the uniformity of the race gap since the screening process (i.e., looking through a large set of resumes) may be quite similar across the variety of jobs we study.[51]

[51] Another explanation could be based on employer stereotyping or categorizing. If employers have coarser stereotypes for African-Americans, many of our results would follow. See Melinda Jones (2002) for the relevant psychology and Mullainathan (2003) for a formalization of the categorization concept.

V. Conclusion

This paper suggests that African-Americans face differential treatment when searching for jobs and this may still be a factor in why they do poorly in the labor market. Job applicants with African-American names get far fewer callbacks for each resume they send out. Equally importantly, applicants with African-American names find it hard to overcome this hurdle in callbacks by improving their observable skills or credentials.

Taken at face value, our results on differential returns to skill have possibly important policy implications. They suggest that training programs alone may not be enough to alleviate the racial gap in labor market outcomes. For training to work, some general-equilibrium force outside the context of our experiment would have to be at play. In fact, if African-Americans recognize how employers reward their skills, they may rationally be less willing than Whites to even participate in these programs.

TABLE A1—FIRST NAMES USED IN EXPERIMENT

White female Name	L(W)/L(B)	Perception White	African-American female Name	L(B)/L(W)	Perception Black
Allison	∞	0.926	Aisha	209	0.97
Anne	∞	0.962	Ebony	∞	0.9
Carrie	∞	0.923	Keisha	116	0.93
Emily	∞	0.925	Kenya	∞	0.967
Jill	∞	0.889	Lakisha	∞	0.967
Laurie	∞	0.963	Latonya	∞	1
Kristen	∞	0.963	Latoya	∞	1
Meredith	∞	0.926	Tamika	284	1
Sarah	∞	0.852	Tanisha	∞	1
	Fraction of all births:			Fraction of all births:	
	3.8 percent			7.1 percent	

White male Name	L(W)/L(B)	Perception White	African-American male Name	L(B)/L(W)	Perception Black
Brad	∞	1	Darnell	∞	0.967
Brendan	∞	0.667	Hakim		0.933
Geoffrey	∞	0.731	Jamal	257	0.967
Greg	∞	1	Jermaine	90.5	1
Brett	∞	0.923	Kareem	∞	0.967
Jay	∞	0.926	Leroy	44.5	0.933
Matthew	∞	0.888	Rasheed	∞	0.931
Neil	∞	0.654	Tremayne	∞	0.897
Todd	∞	0.926	Tyrone	62.5	0.900
	Fraction of all births:			Fraction of all births:	
	1.7 percent			3.1 percent	

Notes: This table tabulates the different first names used in the experiment and their identifiability. The first column reports the likelihood that a baby born with that name (in Massachusetts between 1974 and 1979) is White (or African-American) relative to the likelihood that it is African-American (White). The second column reports the probability that the name was picked as White (or African-American) in an independent field survey of people. The last row for each group of names shows the proportion of all births in that race group that these names account for.

REFERENCES

Aigner, Dennis J. and Cain. Glenn G. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review*, January 1977, *30*(1), pp. 175–87.

Altonji, Joseph G. and Blank, Rebecca M. "Race and Gender in the Labor Markey," in Orley Ashenfelter and David Card, eds., *Handbook of labor economics*, Vol. 30. Amsterdam: North-Holland, 1999, pp. 3143–259.

Arrow, Kenneth, J. "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in labor markets.* Princeton, NJ: Princeton University Press, 1973, pp. 3–33.

_____. "What Has Economics to Say about Racial Discrimination?" *Journal of Economic Perspectives*, Spring 1998, *12*(2), pp. 91–100.

Becker, Gary S. *The economics of discrimination*, 2nd Ed. Chicago: University of Chicago Press, 1961.

Brown, Colin and Gay, Pat. *Racial discrimination 17 years after the act.* London: Policy Studies Institute, 1985.

Cornell, Bradford and Welch, Ivo. "Culture, Information, and Screening Discrimination." *Journal of Political Economy*, June 1996, *104*(3), pp. 542–71.

Council of Economic Advisers. *Changing America: Indicators of social and economic well-being by race and Hispanic origin.* September 1998, http://w3.access.gpo.gov/eop/ca/pdfs/ca.pdf.

Cross, Harry; Kenney, Genevieve; Mell, Jane and Zimmerman, Wendy. *Employer hiring practices: Differential treatment of Hispanic and Anglo job applicants.* Washington, DC: Urban Institute Press, 1990.

Darity, William A., Jr. and Mason, Patrick L. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *Journal of Economic Perspectives*, Spring 1998, *12*(2), pp. 63–90.

Fix, Michael and Turner, Margery A., eds. *A national report card on discrimination in America: The role of testing.* Washington, DC: Urban Institute Press, 1998.

Fryer, Roland and Levitt, Steven. "The Causes and Consequences of Distinctively Black Names." Mimeo, University of Chicago, 2003.

Goldin, Claudia and Rouse, Cecilia. "Orchestrating Impartiality: The Impact of Blind Auditions on Female Musicians." *American Economic Review*, September 2000, *90*(4), pp. 715–41.

Heckman, James J. "Detecting Discrimination." *Journal of Economic Perspectives*, Spring 1998, *12*(2), pp. 101–16.

Heckman, James J.; Lochner, Lance J., and Todd, Petra E. "Fifty Years of Mincer Earnings Regressions." Mimeo, University of Chicago, 2001.

Heckman, James J. and Siegelman, Peter. "The Urban Institute Audit Studies: Their Methods and Findings," in Michael Fix and Raymond J. Struyk, eds., *Clear and convincing evidence: Measurement of discrimination in America.* Lanham, MD: Urban Institute Press, 1992, pp. 187–258.

Holzer, Harry J. "Informal Job Search and Black Youth Unemployment." *American Economic Review*, June 1987, *77*(3), pp. 446–52.

Hubbuck, Jim and Carter, Simon. *Half a chance? A report on job discrimination against young blacks in Nottingham.* London: Commission for Racial Equality, 1980.

James, Franklin and DelCastillo, Steve W. "Measuring Job Discrimination by Private Employers Against Young Black and Hispanic Seeking Entry Level Work in the Denver Metropolitan Area." Mimeo, University of Colorado-Denver, 1991.

Jones, Melinda. *Social psychology of prejudice.* Saddle River, NJ: Pearson Education, 2002.

Jowell, Roger and Prescott-Clark, Patricia. "Racial Discrimination and White-Collar Workers in Britain." *Race*, November 1970, *11*(4), pp. 397–417.

Lundberg, Shelly J. and Starz, Richard. "Private Discrimination and Social Intervention in Competitive Labor Market." *American Economic Review*, June 1983, *73*(3), pp. 340–47.

McIntyre, Shelby J.; Moberg, Dennis J. and Posner, Barry Z. "Discrimination in Recruitment: An Empirical Analysis: Comment." *Industrial and Labor Relations Review*, July 1980, *33*(4), pp. 543–47.

Mullainathan, Sendhil. "Thinking Through Categories." Mimeo, Massachusetts Institute of Technology, 2003.

Neumark, David. "Sex Discrimination in Restaurant Hiring: An Audit Study." *Quarterly Journal of Economics*, August 1996, *111*(3), pp. 915–42.

Newman, Jerry M. "Discrimination in Recruitment: An Empirical Analysis." *Industrial and Labor Relations Review*, October 1978, *32*(1), pp. 15–23.

Nisbett, Richard E. and Cohen, Dov. *The culture of honor: The psychology of violence in the South.* Boulder, CO: Westview Press, 1996.

Phelps, Edmund S. "The Statistical Theory of Racism and Sexism." *American Economic Review*, September 1972, *62*(4), pp. 659–61.

Raphael, Steven; Stoll, Michael A. and Holzer, Harry J. "Are Suburban Firms More Likely to Discriminate against African Americans?" *Journal of Urban Economics*, November 2000, *48*(3), pp. 485–508.

Riach, Peter A. and Rich, Judity. "Testing for Racial Discrimination in the Labour Market." *Cambridge Journal of Economics*, September 1991, *15*(3), pp. 239–56.

Turner, Margery A.; Fix, Michael and Struyk, Raymond J. *Opportunities denied, opportunities diminished: Racial discrimination in hiring.* Washington, DC: Urban Institute Press, 1991.

Weichselbaumer, Doris. "Sexual Orientation Discrimination in Hiring." *Labour Economics*, December 2003, *10*(6), pp. 629–42.

_____. "Is it Sex or Personality? The Impact of Sex-Stereotypes on Discrimination in Applicant Selection." *Eastern Economic Journal*, Spring 2004, *30*(2), pp. 159–86.

What We Know Now About Bias and Intergroup Conflict, the Problem of the Century

Susan T. Fiske[1]

Department of Psychology, Princeton University, Princeton, New Jersey

Abstract

After nearly a century's study, what do psychologists now know about intergroup bias and conflict? Most people reveal unconscious, subtle biases, which are relatively automatic, cool, indirect, ambiguous, and ambivalent. Subtle biases underlie ordinary discrimination: comfort with one's own in-group, plus exclusion and avoidance of out-groups. Such biases result from internal conflict between cultural ideals and cultural biases. A small minority of people, extremists, do harbor blatant biases that are more conscious, hot, direct, and unambiguous. Blatant biases underlie aggression, including hate crimes. Such biases result from perceived intergroup conflict over economics and values, in a world perceived to be hierarchical and dangerous. Reduction of both subtle and blatant bias results from education, economic opportunity, and constructive intergroup contact.

Keywords

bias; stereotyping; prejudice; discrimination; intergroup conflict

People typically seek other people who are similar to themselves, being comfortable with others they perceive as members of their own in-group. From comfort follows, at best, neglect of people from out-groups and, at worst, murderous hostility toward out-groups perceived as threatening the in-group. Biases do vary by degree, and the psychologies of moderate and extreme biases differ considerably. Well-intentioned moderates reveal bias more subtle than the rants and rampages of extremists. By some counts, 80% of Western democratic populations intend benign intergroup relations but display subtle biases. In contrast, blatantly biased extremists are completely out-front. Although estimated to be a minority (perhaps 10%), they are salient, vocal, and dangerous.

After nearly a century's study, social psychology knows a lot about both forms of bias. Stereotyping, prejudice, and discrimination reflect, respectively, people's cognitive, affective, and behavioral reactions to people from other groups (Fiske, 1998). All constitute *bias*, reacting to a person on the basis of perceived membership in a single human category, ignoring other category memberships and other personal attributes. Bias is thus a narrow, potentially erroneous reaction, compared with individuated impressions formed from personal details.

SUBTLE BIAS AMONG WELL-INTENTIONED MODERATES

Automatic, Unconscious, and Unintentional

The big news from two recent decades of research: Bias is most often

Copyright © 2002 American Psychological Society

underground (Dovidio & Gaertner, 1986). First data showed that even among relatively unprejudiced people, racial category labels automatically prime (increase the accessibility of) stereotypes; scores of studies now support the essential automaticity of stereotypes (Fiske, 1998, 2000; Macrae & Bodenhausen, 2000). For example, even when outgroup category labels are subliminally presented (i.e., presented too quickly to be consciously perceived), they activate stereotypic associations. In a more affective vein, out-group cues (such as faces or names) easily activate negative evaluative terms. Relatedly, brain imaging shows activation of the amygdala in response to out-group faces; because the amygdala is the center of fear and anxiety in the brain, its activation in response to out-groups is consistent with primitive emotional prejudices. Furthermore, automatic activation of outgroup categories leads to behavior stereotypically associated with that group. For example, young people primed with the category "elderly" (vs. a neutral one) walk and respond more slowly; Whites primed with Black faces (vs. White ones) respond in a more hostile way than they normally do. Such responses create self-fulfilling prophecies in intergroup biases. Brain imaging and automatic behavior form the cutting edges of work on automatic biases.

Automatic reactions to outgroup members matter in everyday behavior. Awkward social interactions, embarrassing slips of the tongue, unchecked assumptions, stereotypic judgments, and spontaneous neglect all exemplify the mundane automaticity of bias, which creates a subtly hostile environment for out-group members. The apparent automaticity of routine biases corroborates Allport's (1954) provocative early insights about the inevitability of categorization. Automaticity also shocks well-intentioned people who assume that both their own and other people's prejudice must be conscious and controllable.

All is not lost for the well-intentioned. Category activation is not *unconditionally* automatic. Although people can instantly identify another person's category membership (especially gender, race, and age), they may not always activate associated stereotypes. For example, sufficient mental overload blocks activation. People's long-term attitudes also have a moderating influence: Chronically low levels of prejudice can attenuate the activation of stereotypes. Temporary goals matter, too: Category activation depends on short-term motivations, including immediate threats to self-esteem and focused efforts toward accurate understanding.

Promising as they are, findings indicating that biases are automatic, unconscious, and unintentional remain controversial. For example, the ease of category activation differs depending on the nature of the stimuli: Activation is easy when people encounter verbal labels, harder when they encounter photographs, and hardest when they encounter real people. Some researchers believe that social categories inevitably activate associated biases, whereas others believe activation depends entirely on short-term goals and long-term individual differences (Devine, 2001).

Whether bias is conditionally or unconditionally automatic, less prejudiced perceivers still can compensate for their automatic associations with subsequent conscious effort. If category activation is conditionally automatic, then people may be able to inhibit it in the first place. In either case, motivation matters.

Moreover, even if people do activate biases associated with a category, they may not apply (or use) those biases. For example, once the category is activated, other information may be consistent or inconsistent with it, and perceivers have to decide what to do about the conflicting information. Inconsistency resolution and subsequent individuation of the other person require mental resources, which are allocated according to the perceiver's motivation and capacity. Overriding category use depends on metacognitive decisions (thinking about one's thinking) and higher-level executive functions (controlling one's thinking), not just brute attentional capacity. Other influences on whether activated categories are used go beyond the perceiver's motivation and capacity: For example, category use depends on the stimuli (general group-level abstractions encourage assimilation toward the stereotype, whereas individual exemplars encourage contrast away from it). Category use increases when the perceiver's personal theory holds that people's dispositions are fixed entities, rather than flexible states. Psychologists continue to debate the boundaries of automaticity.

Inhibition of both category activation and category application challenges even the most determined moderate. Direct suppression sometimes causes only a rebound of the forbidden biases. Depending on cognitive capacity, practice, age, and motivation, people can inhibit many effects of social categories on their thinking, feeling, and behaving. Indeed, when people adopt goals encouraging them to treat others as unique individuals or not at all as social objects, they no longer show even amygdala activation in response to faces from races other than their own (Wheeler & Fiske, 2001). The take-home message: Bias is more automatic than people think, but less automatic than psychologists thought.

Cool, Indirect, and Ambiguous

The biases of the moderate, well-intentioned majority not only live underground, they also wear

camouflage. Consistent with people's biases reflecting in-group comfort at least as much as out-group discomfort, bias often consists of withholding positive emotions from out-groups. Moderates rarely express open hostility toward out-groups, but they may withhold basic liking and respect; hence, their responses represent cool neglect. People more rapidly assign positive attributes to the in-group than the out-group, but often show at best weak differences in assigning negative attributes. People withhold rewards from out-groups, relative to the in-group, reflecting favoritism (Brewer & Brown, 1998). But they rarely punish or derogate the out-group. The typical damage is relative.

Moderate biases are indirect, relying on norms for appropriate responses. If norms allow biases, they flourish. That is, biases most often appear when people have unprejudiced excuses. When contact or helping is discretionary, for example, if some people neglect out-group members, then most people do. If the out-group member behaves poorly, providing an excuse for prejudice, the resulting exclusion is more swift and sure than for a comparable in-group member. Biases also appear in political policy preferences for which one might have principled reasons (excuses), but regarding which one also just happens to have a series of opinions that all disadvantage the out-group relative to the in-group. Excuses for bias fulfill the social norm requiring rational, fair judgments, but empirically controlled comparisons reveal greater bias toward out-group members than toward comparable in-group members. Researchers continue to debate the meaning of these biases.

People also engage in attributional tricks that discourage sympathy by blaming the out-group for their own unfortunate outcomes: Members of the out-group should try harder, but at the same time they should not push themselves where they are not wanted (Catch 22). The blame goes further. Although the in-group might be excused for its failures (extenuating circumstances), the out-group brought it on themselves (unfortunate dispositions). People often attribute the out-group's perceived failings to their essence: Innate, inherent, enduring attributes, perhaps biological, especially genetic, define category distinctiveness.

In making sense of out-group members, people exaggerate cultural differences (in ability, language, religious beliefs, and sexual practices). The mere fact of categorizing into in-group "us" and out-group "them" exaggerates intercategory differences and diminishes intracategory differences: "Out-group members all are alike and different from us, besides." In short, moderates' bias is cool, indirect, and ambiguous.

Ambivalent and Mixed

Besides being underground and camouflaged, moderate biases are complex. Ambivalent racism entails, for moderate Whites, mixed "pro-Black" pity and anti-Black resentment, which tips over to a predominantly positive or negative response, depending on circumstances. Ambivalent sexism is another example, demonstrating two correlated dimensions that differentiate hostile sexism (toward nontraditional women) and subjectively benevolent sexism (toward traditional women). In both cases, ambivalence indicates mixed forms of prejudice more subtle than unmitigated hostility.

Mixed biases turn out to be the rule, rather than the exception. Although various out-groups all are classified as "them," they form clusters (see Fig. 1). Some elicit less respect than others, and some elicit less liking than others. Not only is the bias of well-intentioned moderates of the cool variety (withholding the positive, rather than assigning the negative), but it is not even uniformly lacking in positive views. Specifically, some out-groups (Asians, Jews, career women, Black professionals, rich people) are envied and respected for their perceived competence and high status, but they are resented and disliked as lacking in warmth because they compete with the in-group. Other out-groups (older people, disabled people, housewives) are pitied and disrespected for their perceived incompetence and low status, but they are nurtured and liked as warm because they do not threaten the in-group. Only a few out-groups (primarily homeless and poor people of any race) receive contempt, both dislike and disrespect, because they are seen as simultaneously low status and exploiting the in-group.

Ambivalent, mixed biases justify the status quo. Subordinated, pitied groups have an incentive to cooperate because they receive care, in return for not challenging the hierarchy. Conversely, dominant groups use subordinated groups to maintain their own relative advantage. Envied, competitive groups have an incentive to support the system because they are perceived to be succeeding, even if they are socially excluded by the culturally dominant group. For dominant groups, respecting envied groups acknowledges the ground rules for competition (which favor them also), but disliking those groups justifies social exclusion.

Moderate Biases Lead to Exclusion

Subtle biases motivate personal interactions that reek of discomfort and anxiety. Nonverbal indicators (distance, posture, voice tone) and people's own reports of their feelings all reveal intergroup interactions that are anything but smooth,

Copyright © 2002 American Psychological Society

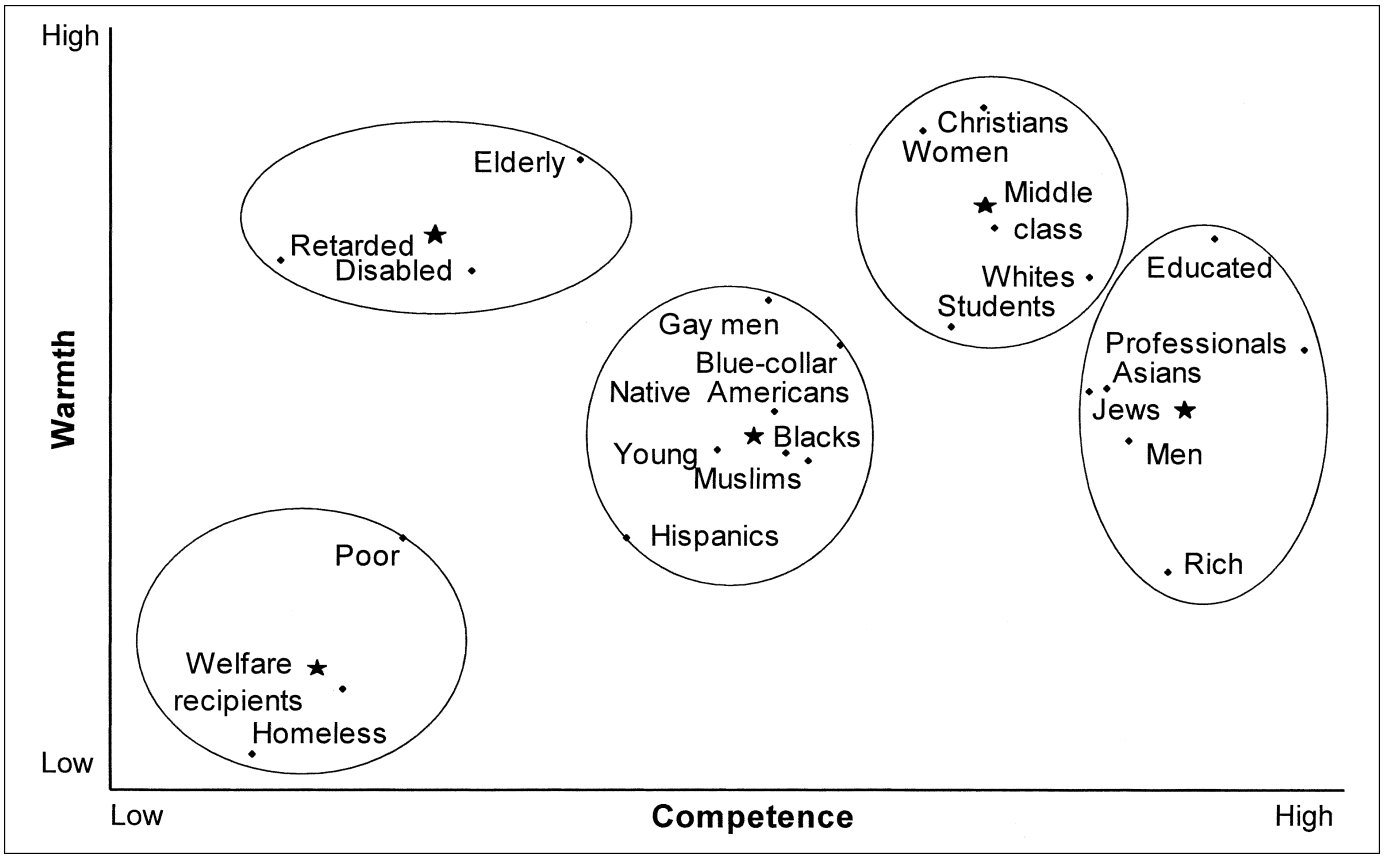


Fig. 1. Five-cluster solution showing the perceived distribution of American social groups, according to perceived competence and warmth (Fiske, Cuddy, Glick, & Xu, 2002, Study 2). Copyright by the American Psychological Association. Reprinted with permission.

mostly because of inexperience with the out-group.

Moreover, people mentally and behaviorally confirm their biased expectations, leading both groups to maintain their distance. *Self-fulfilling prophecy*, *expectancy effects*, and *behavioral confirmation* all name related phenomena whereby biased perceivers bring about the very behavior they anticipate, usually negative. These interpersonal processes result in subsequent avoidance, whenever people can choose the company they keep. Discretionary intergroup contact is minimized.

Furthermore, exclusion and avoidance extend to employment, housing, education, and justice that tend to favor the in-group and disadvantage the out-group. Ample evidence indicates that relatively automatic, cool, indirect, ambiguous, and ambivalent biases permit allocation of resources to maintain the in-group's advantage.

How Do Moderate Biases Originate?

Subtle prejudice comes from people's internal conflict between ideals and biases, both acquired from the culture. Direct, personal experience with out-group members may be limited. Given substantial de facto residential and occupational segregation, people lack experience in constructive intergroup interactions. Cultural media, then, supply most information about out-groups, so people easily develop unconscious associations and feelings that reinforce bias.

Simultaneously, contemporary Western ideals encourage tolerance of most out-groups. Complying with modern antiprejudice ideals requires conscious endorsement of egalitarian norms against prejudice. And moderates do endorse antiprejudice values. The upshot is a conflict between relatively implicit, unconscious biases and explicit, conscious ideals to be unprejudiced. The resulting prejudices are subtle, modern, and aversive to the people holding them.

BLATANT BIAS AMONG ILL-INTENTIONED EXTREMISTS

Hot, Direct, Unambiguous, and Conscious

In contrast to well-intentioned moderates, extremists openly resent

out-groups and reject any possibility of intimacy with them (Pettigrew, 1998b). They resent out-groups—whether racial, cultural, gender, or sexual—as holding jobs that in-group members should have and (paradoxically) living on welfare unnecessarily. They believe that out-groups and the in-group can never be comfortable together. Extremists are particularly upset by intergroup intimacy. They report that they would be bothered by having a mixed grandchild, that they are unwilling to have sexual relations with out-group members, and that they are unwilling to have an out-group boss.

Extreme biases run in packs; people biased against one out-group tend to be biased against others. People's differing levels of ethnocentrism are reliably measured by old-fashioned prejudice scales that assess self-reported attitudes toward racial, ethnic, gender, and sexual out-groups.

Extreme Biases Underlie Aggression

The result is as simple as it is horrible. Because of hot, direct, unambiguous prejudices, extremists advocate segregation, containment, and even elimination of out-groups. Strong forms of bias correlate with approval of racist movements. Hate-crime perpetrators and participants in ethnic violence, not surprisingly, endorse attitudes (prejudices and stereotypes) that fit extreme forms of discrimination.

Aggression has two main goals: preserving hierarchies and preserving values perceived to be traditional. People with blatant prejudices often approve aggression to maintain the status quo, viewing current group hierarchies as inevitable and desirable. Highly prejudiced people gravitate toward jobs that enhance group hierarchy and defend the status quo (e.g., they tend to be police officers rather than social workers and businesspeople rather than educators). Blatant prejudice may also lead to self-righteous aggression against nonconformers and other people who threaten core values. If out-groups deviate and threaten traditional values, they become legitimate targets of aggression.

How Do Extreme Biases Originate?

Whether extremists are domestic or international, they endanger those they hate. People become biased extremists because they perceive threats to their in-group. Thus, extreme bias parallels the in-group favoritism of biased moderates, who also protect the in-group. Differences lie in the perceived nature and degree of threat.

Threat to economic standing has long been implicated in intergroup bias. Although still controversial, the most convincing but counterintuitive lesson here is that personal economic deprivation is not in fact the culprit. The state of people's own wallets does not motivate their degree of prejudice. Instead, the most reliable indicator is perceived threat to one's in-group. Group threat (e.g., high local unemployment) correlates with extreme biases against out-groups perceived to be responsible. The causal sequence seems to run from subjective social class to perceived group deprivation to prejudice.

Perceived threat to in-group economic status correlates with worldviews that reinforce a zero-sum, dog-eat-dog perspective. Tough-minded competition is perceived to reflect the state of intergroup relations. Economic conservatism results. Overall, blatant prejudice correlates with high social dominance orientation (Sidanius & Pratto, 1999), that is, endorsing views that superior groups should dominate inferior groups, that force may be necessary to maintain this dominance, and that group equality is neither desirable nor realistic.

Perceived threat to traditional values is the other prong of blatant bias. Extremists view the world as dangerous, with established authority and conventions in collapse. Social conservatism correlates with perceived threats to traditional values, and also with extreme bias (Altemeyer, 1996). Extremists move in tight ethnic circles and endorse right-wing authoritarian views: old-fashioned values, censorship, mighty leaders who fight evil, and suppression of troublemakers, freethinkers, women, and homosexuals.

The background of people who become extremists features limited intergroup contact—few out-group neighbors, acquaintances, and friends. Nor do extremists value such contact. They also tend to be less educated than moderates, for reasons not fully clear, although one might speculate that a liberal education broadens people's appreciation of different values.

WHAT REDUCES BIAS?

Given subtle biases that are unconscious and indirect, change is a challenge, resisting frontal assault. Similarly, given blatant biases rooted in perceived threat to group interests and core values, direct confrontation will likely fail again. Instead, more nuanced means do work.

Education does help. Economic opportunity does help. Moreover, for decades, social psychologists have studied the positive effects of constructive intergroup contact that increases mutual appreciation (Pettigrew, 1998a): When contact features (a) equal status within the immediate setting, (b) shared

Copyright © 2002 American Psychological Society

goals, (c) cooperation in pursuit of those goals, and (d) authorities' support, it provides a basis for intergroup friendship. Genuine intergroup friendships demonstrably do reduce stereotyping, prejudice, and discrimination of whatever sort.

WHERE NOW?

Much is known, but much remains to be learned. Promising lines of research range from imaging brain activity beyond the amygdala, to specifying intergroup emotions beyond mere antipathy, to explaining stereotype content beyond mere lists of negative traits, to predicting discrimination in all its guises, to assessing people's control over their own seemingly automatic reactions (Fiske, 2000). Bias researchers will not be unemployed any time soon.

Recommended Reading

Duckitt, J. (1992). *The social psychology of prejudice*. New York: Praeger.

Eberhardt, J.L., & Fiske, S.T. (Eds.). (1998). *Confronting racism: The problem and the response*. Thousand Oaks, CA: Sage.

Oskamp, S. (Ed.). (2000). *Reducing prejudice and discrimination*. Mahwah, NJ: Erlbaum.

Note

1. Address correspondence to Susan T. Fiske, Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544-1010; e-mail: sfiske@princeton.edu.

References

Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Altemeyer, B. (1996). *The authoritarian specter*. Cambridge, MA: Harvard University Press.

Brewer, M.B., & Brown, R.J. (1998). Intergroup relations. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 554–594). New York: McGraw-Hill.

Devine, P.G. (2001). Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section. *Journal of Personality and Social Psychology*, *81*, 757–759.

Dovidio, J.F., & Gaertner, S.L. (Eds.). (1986). *Prejudice, discrimination, and racism*. San Diego, CA: Academic Press.

Fiske, S.T. (1998). Stereotyping, prejudice, and discrimination. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357–411). New York: McGraw-Hill.

Fiske, S.T. (2000). Stereotyping, prejudice, and discrimination at the seam between the centuries: Evolution, culture, mind and brain. *European Journal of Social Psychology*, *30*, 299–322.

Fiske, S.T., Cuddy, A.J., 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.

Macrae, C.N., & Bodenhausen, G.V. (2000). Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.

Pettigrew, T.F. (1998a). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85.

Pettigrew, T.F. (1998b). Reactions toward the new minorities of western Europe. *Annual Review of Sociology*, *24*, 77–103.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.

Wheeler, M.E., & Fiske, S.T. (2001, November). *fMRI study of three cognitive tasks that differentially modulate stereotype accessibility and human amygdala response to racial outgroup faces*. Poster presented at the annual meeting of the Society for Neuroscience, San Diego, CA.

A Threat in the Air

How Stereotypes Shape Intellectual Identity and Performance

Claude M. Steele
Stanford University

A general theory of domain identification is used to describe achievement barriers still faced by women in advanced quantitative areas and by African Americans in school. The theory assumes that sustained school success requires identification with school and its subdomains; that societal pressures on these groups (e.g., economic disadvantage, gender roles) can frustrate this identification; and that in school domains where these groups are negatively stereotyped, those who have become domain identified face the further barrier of stereotype threat, *the threat that others' judgments or their own actions will negatively stereotype them in the domain. Research shows that this threat dramatically depresses the standardized test performance of women and African Americans who are in the academic vanguard of their groups (offering a new interpretation of group differences in standardized test performance), that it causes disidentification with school, and that practices that reduce this threat can reduce these negative effects.*

From an observer's standpoint, the situations of a boy and a girl in a math classroom or of a Black student and a White student in any classroom are essentially the same. The teacher is the same; the textbooks are the same; and in better classrooms, these students are treated the same. Is it possible, then, that they could still experience the classroom differently, so differently in fact as to significantly affect their performance and achievement there? This is the central question of this article, and in seeking an answer, it has both a practical and a theoretical focus. The practical focus is on the perhaps obvious need to better understand the processes that can hamper a group's school performance and on what can be done to improve that performance. The theoretical focus is on how societal stereotypes about groups can influence the intellectual functioning and identity development of individual group members. To show the generality of these processes and their relevance to important outcomes, this theory is applied to two groups: African Americans, who must contend with negative stereotypes about their abilities in many scholastic domains, and women, who must do so primarily in math and the physical sciences. In trying to understand the schooling outcomes of these two groups, the theory has a distinct perspective, that of viewing people, in Sartre's (1946/1965) words, as "first of all beings in a situation" such that if one wants to understand them, one "must inquire first into the situation surrounding [them]" (p. 60).

The theory begins with an assumption: that to sustain school success one must be identified with school achievement in the sense of its being a part of one's self-definition, a personal identity to which one is self-evaluatively accountable. This accountability—that good self-feelings depend in some part on good achievement—translates into sustained achievement motivation. For such an identification to form, this reasoning continues, one must perceive good prospects in the domain, that is, that one has the interests, skills, resources, and opportunities to prosper there, as well as that one belongs there, in the sense of being accepted and valued in the domain. If this relationship to schooling does not form or gets broken, achievement may suffer. Thus, in trying to understand what imperils achievement among women and African Americans, this logic points to a basic question: What in the experience of these groups might frustrate their identification with all or certain aspects of school achievement?

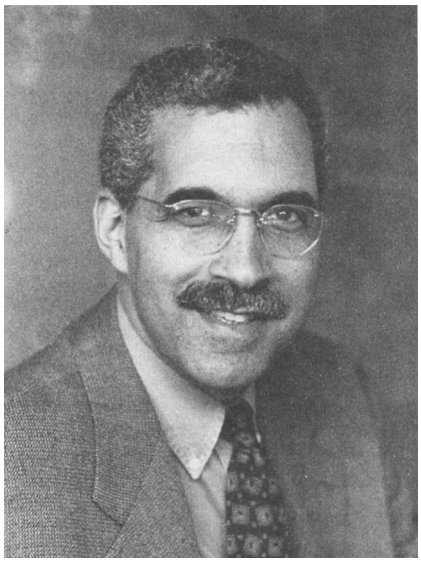
One must surely turn first to social structure: limits on educational access that have been imposed on these groups by socioeconomic disadvantage, segregating social practices, and restrictive cultural orientations, limits of both historical and ongoing effect. By diminishing one's educational prospects, these limitations (e.g., inadequate resources, few role models, preparational disadvantages) should make it more difficult to identify with academic domains. To continue in math, for example, a woman might have to buck the low expectations of teachers, family, and societal gender roles in which math is seen as unfeminine as well as anticipate spending her entire professional life in a male-dominated world. These

Editor's note. Cheryl B. Travis served as action editor for this article.

Author's note. The research reported in this article was supported by National Institutes of Health Grant MH51977, Russell Sage Foundation Grant 879.304, and Spencer Foundation and James S. McDonnell Foundation postdoctoral fellowships. Completion of the research was aided by the Center for Advanced Study in the Behavioral Sciences.

Correspondence concerning this article should be addressed to Claude M. Steele, Department of Psychology, Jordan Hall, Building 420, Stanford University, Stanford, CA 94305. Electronic mail may be sent via Internet to steele@psych.stanford.edu.

Copyright 1997 by the American Psychological Association, Inc. 0003-066X/97/$2.00
Vol. 52, No. 6, 613–629



Claude M. Steele
Copyright by L. A. Cicero, Stanford News Service.

realities, imposed on her by societal structure, could so reduce her sense of good prospects in math as to make identifying with it difficult.

But this article focuses on a further barrier, one that has its effect on the already identified, those members of these groups who, having survived structural obstacles, have achieved identification with the domain (of the present groups, school-identified African Americans and math-identified women). It is the social–psychological threat that arises when one is in a situation or doing something for which a negative stereotype about one's group applies. This predicament threatens one with being negatively stereotyped, with being judged or treated stereotypically, or with the prospect of conforming to the stereotype. Called *stereotype threat,* it is a situational threat—a threat in the air—that, in general form, can affect the members of any group about whom a negative stereotype exists (e.g., skateboarders, older adults, White men, gang members). Where bad stereotypes about these groups apply, members of these groups can fear being reduced to that stereotype. And for those who identify with the domain to which the stereotype is relevant, this predicament can be self-threatening.

Negative stereotypes about women and African Americans bear on important academic abilities. Thus, for members of these groups who are identified with domains in which these stereotypes apply, the threat of these stereotypes can be sharply felt and, in several ways, hampers their achievement.

First, if the threat is experienced in the midst of a domain performance—classroom presentation or test-taking, for example—the emotional reaction it causes could directly interfere with performance. My colleagues and I (Spencer, Steele, & Quinn, 1997; C. M. Steele & Aronson, 1995) have tested this possibility with women taking standardized math tests and African Americans taking standardized verbal tests. Second, when this threat becomes chronic in a situation, as for the woman who spends considerable time in a competitive, male-oriented math environment, it can pressure *disidentification,* a reconceptualization of the self and of one's values so as to remove the domain as a self-identity, as a basis of self-evaluation. Disidentification offers the retreat of not caring about the domain in relation to the self. But as it protects in this way, it can undermine sustained motivation in the domain, an adaptation that can be costly when the domain is as important as schooling.

Stereotype threat is especially frustrating because, at each level of schooling, it affects the vanguard of these groups, those with the skills and self-confidence to have identified with the domain. Ironically, their susceptibility to this threat derives not from internal doubts about their ability (e.g., their internalization of the stereotype) but from their identification with the domain and the resulting concern they have about being stereotyped in it. (This argument has the hopeful implication that to improve the domain performance of these students, one should focus on the feasible task of lifting this situational threat rather than on altering their internal psychology.) Yet, as schooling progresses and the obstacles of structure and stereotype threat take their cumulative toll, more of this vanguard will likely be pressured into the ranks of the unidentified. These students, by not caring about the domain vis-à-vis the self, are likely to underperform in it regardless of whether they are stereotype threatened there. Thus, although the identified among these groups are likely to underperform only under stereotype threat, the unidentified (casualties of sociocultural disadvantage or prior internalization of stereotype threat) are likely to underperform and not persist in the domain even when stereotype threat has been removed.

In these ways, then, the present analysis sees social structure and stereotypes as shaping the academic identities and performance outcomes of large segments of society. But first, for the two groups under consideration, what are these outcomes?

As is much discussed, these outcomes are in a crisis state for African Americans. Although Black students begin school with standardized test scores that are not too far behind those of their White counterparts, almost immediately a gap begins to appear (e.g., Alexander & Entwistle, 1988; Burton & Jones, 1982; Coleman et al., 1966) that, by the sixth grade in most school districts, is two full grade levels (Gerard, 1983). There have been encouraging increases in the number of African Americans completing high school or its equivalence in recent years: 77% for Black students versus 83% for White students (American Council on Education, 1995–1996). And there have been modest advances in the number of African American high school graduates enrolling in college, although these have not been as substantial as in other groups (American Council on Education, 1995–1996). Perhaps most discouraging has been the high dropout rate for African American college students: Those who do not finish college within six years is 62%, com-

pared with a national dropout rate of 41% (American Council on Education, 1995–1996). And there is evidence of lower grade performance among those who do graduate of, on average, two thirds of a letter grade lower than those of other graduating students (Nettles, 1988). On predominantly White campuses, Black students are also underrepresented in math and the natural sciences. Although historically Black colleges and universities now enroll only 17% of the nation's Black college students, they produce 42% of all Black BS degrees in natural science (Culotta & Gibbons, 1992). At the graduate level, although Black women have recently shown modest gains in PhDs received, the number awarded to Black men has declined over the past decade more than for any other subgroup in society (American Council on Education, 1995–1996).

Women clearly thrive in many areas of schooling. But in math, engineering, and the physical sciences, they often endure lesser outcomes than men. In a meta-analysis involving over 3 million participants, Hyde, Fennema, and Lamon (1990), for example, found that through elementary and middle school, there are virtually no differences between boys and girls in performance on standardized math tests but that a trend toward men doing better steadily increases from high school (SD = .29) through college (SD = .41) and into adulthood (SD = .59). And, as their college careers begin, women leave these fields at a rate two and a half times that of men (Hewitt & Seymour, 1991). Although White women constitute 43% of the U.S. population, they earn only 22% of the BS degrees and 13% of the PhDs and occupy only 10% of the jobs in physical science, math, and engineering, where they earn only 75% of the salary paid to men (Hewitt & Seymour, 1991).

These inequities have compelled explanations ranging from the sociocultural to the genetic. In the case of African Americans, for example, past and ongoing socioeconomic disadvantage, cultural orientations (e.g., Ogbu, 1986), and genetic differences (e.g., Herrnstein & Murray, 1994; Jensen, 1969) have all been proposed as factors that, through singular and accumulated effect, could undermine their performance. In the case of women's performance in math and the physical sciences, there are parallel arguments: structural and cultural gender role constraints that shunt women away from these areas; culturally rooted expectations (e.g., Eccles, 1987; Eccles-Parsons et al., 1983); and, again, genetic limitations (Benbow & Stanley, 1980, 1983). But, like crumbs along the forest floor, several findings lead away from these analyses as fully sufficient.

For one thing, minority student achievement gaps persist even in the middle and upper socioeconomic classes. Using data from the Coleman report (Coleman et al., 1966) and a more recent College Board study of Scholastic Assessment Test (SAT) scores, Miller (1995, 1996) found that the gaps in academic performance (grades as well as standardized test scores) between Whites and non-Asian minorities (e.g., African Americans, Hispanics, and Native Americans) were as large, or larger, in the upper and middle classes (as measured by parental education and occupation) than in the lower classes. Group differences in socioeconomic status (SES), then, cannot fully explain group differences in academic performance.

Another point is that these differences are not even fully explained by group differences in skills. This is shown in the well-known *overprediction* or *underperformance* phenomenon of the test bias literature. Overprediction occurs when, at each level of performance on a test of preparation for some level of schooling (e.g, the SAT), students from one group wind up achieving less—getting lower college grades, for example—than other students with the same beginning scores. In this sense, the test scores of the low-performing group overpredict how well they will actually achieve, or, stated another way, the low-performing group underperforms in relation to the test's prediction. But the point here is that because the students at each test-score level have comparable initial skills, the lower eventual performance of one group must be due to something other than skill deficits they brought with them.

In the case of African Americans, overprediction across the academic spectrum has been so reliably observed as to be almost a lawful phenomenon in American society (e.g., Jensen, 1980; Vars & Bowen, 1997). Perhaps the most extensive single demonstration of it comes from a recent Educational Testing Service study (Ramist, Lewis, & McCamley-Jenkins, 1994) that examined the predictiveness of the SAT on 38 representative college and university campuses. As is typically the case, the study found that the predictive validity to the SAT—its correlation with subsequent grades—was as good for African American, Hispanic, and Native American students as for White and Asian students. But for the three non-Asian minority groups, there was sizable overprediction (underperformance) in virtually all academic areas. That is, at each level of preparation as measured by the SAT, something further depressed the grades of these groups once they arrived on campus.

As important, the same study found evidence of SAT overprediction for female students (i.e., women performing less well than men at comparable SAT levels) in technical and physical science courses such as engineering, economics, and computer science but not in nontechnical areas such as English. It is interesting though that women in this study were not overpredicted in math per se, a seeming exception to this pattern. The overprediction of women's college math performance has generally been unreliable, with some studies showing it (e.g., Benbow & Arjmand, 1990; Levin & Wyckoff, 1988; Lovely, 1987; Ware, Steckler, & Leserman, 1985) and others not (e.g., Adelman, 1991; DeBoer, 1984; Ware & Dill, 1986). However, a recent study (Strenta, Elliott, Adair, Scott, & Matier, 1993) involving over 5,000 students at four prestigious northeastern colleges identified a pattern of effects that suggests why these different results

occur: Underperformance reliably occurred among women who were talented in math and science and who, perhaps for that reason, took courses in these areas that were intended for majors, whereas it did not occur among women with less math and science preparation who took courses in these areas intended for nonmajors. Thus, women may be reliably overpredicted in math and the physical sciences, just as Black students are more generally, but only when the curriculum is more advanced and only among women who are more identified with the domain. Among this vanguard, though, something other than skill deficits depresses their performance. What are these further processes?

Social and Stereotype Structure as Obstacles to Achievement Identification

The proposed answer is that at least one of these processes is a set of social psychological phenomena that obstructs these groups' identification with domains of schooling.[1] I turn first to school identification.

Academic Identification

As noted, this analysis assumes that sustained school achievement depends, most centrally, on identifying with school, that is, forming a relationship between oneself and the domains of schooling such that one's self-regard significantly depends on achievement in those domains. Extrinsic rewards such as better career outcomes, personal security, parental exhortation, and so on, can also motivate school achievement. But it is presumed that sustaining motivation through the ebb and flow of these other rewards requires school identification. How, then, is this identification formed?

Not a great deal is known about the process. But several models (e.g., Schlenker & Weigold, 1989; C. M. Steele, 1988; Tesser, 1988) share an implicit reasoning, the first assumption of which is that people need positive self-regard, a self-perception of "adaptive and moral adequacy" (C. M. Steele, 1988, p. 289). Then, the argument goes, identification with a given domain of life depends, in large part, on the self-evaluative prospects it offers. James (1890/1950) described the development of the self as a process of picking from the many, often incompatible, possible selves, those "on which to stake one's salvation" (p. 310). This choice and the assessment of prospects that goes into it are, of course, multifaceted: Are the rewards of the domain attractive or important? Is an adequate opportunity structure available? Do I have the requisite skills, talents, and interests? Have others like me succeeded in the domain? Will I be seen as belonging in the domain? Will I be prejudiced against in the domain? Can I envision wanting what this domain has to offer? and so on. Some of these assessments undergird a sense of efficacy in the domain (e.g., Bandura, 1977, 1986). Others have to do with the rewards, importance, and attractiveness of the domain itself. And still others have to do with the feasibility and receptiveness of the domain. The point here is that students tacitly assess their prospects in school and its subdomains, and, roughly speaking, their identifications follow these assessments: increasing when they are favorable and decreasing when they are unfavorable. As for the two groups under consideration, then, this analysis suggests that something systematically downgrades their assessments of, and thus their identification with, critical domains of schooling.

Threats to Academic Identification

Structural and cultural threats. Both groups have endured and continue to endure sociocultural influences that could have such effects. Among the most replicable facts in the schooling literature is that SES is strongly related to school success and cognitive performance (e.g., Coleman et al., 1966; Miller, 1996). And because African Americans have long been disproportionately represented in lower socioeconomic classes, this factor surely contributes to their achievement patterns in school, both through the material limitations associated with lower SES (poor schools, lack of resources for school persistence, etc.) and through the ability of these limitations, by downgrading school-related prospects, to undermine identification with school. And beyond socioeconomic structure, there are cultural patterns within these groups or in the relation between these groups and the larger society that may also frustrate their identification with school or some part of it, for example, Ogbu's (1986) notion of a lower-class Black culture that is "oppositional" to school achievement or traditional feminine gender roles that eschew math-related fields (e.g., Eccles-Parsons et al., 1983; Linn, 1994).

Stereotype threat. Beyond these threats, waiting for those in these groups who have identified with school, is yet another threat to their identification, more subtle perhaps but nonetheless profound: that of stereotype threat. I define it as follows: the event of a negative stereotype about a group to which one belongs becoming self-relevant, usually as a plausible interpretation for something one is doing, for an experience one is having, or for a situation one is in, that has relevance to one's self-definition. It happens when one is in the *field* of the stereotype, what Cross (1991) called a "spotlight anxiety" (p. 195), such that one can be judged or treated in terms of a racial stereotype. Many analysts have referred to this predicament and the pressure it causes (e.g., Allport, 1954; Carter, 1991; Cose, 1993; Goffman, 1963; Howard & Hammond, 1985; E. E. Jones et al., 1984; Sartre, 1946/1965; C. M. Steele, 1975; C. M. Steele &

[1] Other factors may also contribute. For example, there are persistent reports of women and minorities being treated differently in the classroom and in other aspects of schooling (e.g., Hewitt & Seymour, 1991). This treatment includes both the "chilly-climate" sins of omission—the failure to call on them in class or to recognize and encourage their talents, and so on—and, in the case of low-income minorities, sins of commission—disproportionate expulsion from school, assignment to special education classes, and administration or corporal punishment ("National Coalition of Advocates for Students Report," 1988).

Aronson, 1995; S. Steele, 1990). The present definition stresses that for a negative stereotype to be threatening, it must be self-relevant. Then, the situational contingency it establishes—the possibility of conforming to the stereotype or of being treated and judged in terms of it—becomes self-threatening. It means that one could be limited or diminished in a domain that is self-definitional. For students from groups in which abilities are negatively stereotyped in all or some school domains and yet who remain identified with those domains, this threat may be keenly felt, felt enough, I argue, to become a further barrier to their identification with the domain.

There is, however, a more standard explanation of how negative stereotypes affect their targets. Beginning with Freud (as cited in Brill, 1938) in psychology and Cooley (1956) and Mead (1934) in sociology, treatises on the experience of oppression have depicted a fairly standard sequence of events: Through long exposure to negative stereotypes about their group, members of prejudiced-against groups often internalize the stereotypes, and the resulting sense of inadequacy becomes part of their personality (e.g., Allport, 1954; Bettelheim, 1943; Clark, 1965; Grier & Coobs, 1968; Erikson, 1956; Fanon, 1952/1967; Kardiner & Ovesey, 1951; Lewin, 1941).

In recent years, the tone of this argument has constructively lightened, replacing the notion of a broad self-hatred with the idea of an inferiority anxiety or low expectations and suggesting how situational factors contribute to this experience. S. Steele's (1990) essays on *racial vulnerability* (i.e., a vulnerability of both Blacks and Whites that stems, in part, from the situational pressures of reputations about their groups) offered an example. This work depicts the workings of this anxiety among African Americans in an interconnected set of ideas: *integration shock* that, like Goffman (1963), points to settings that integrate Blacks and Whites as particularly anxiety arousing; *objective correlatives* or race-related situational cues that can trigger this anxiety; and the inherent sense of risk, stemming from an internalized *inferiority anxiety* and from a *myth of inferiority* pervading integrated settings, of being judged inferior or of confirming one's own feared inferiority. Howard and Hammond (1985) earlier made this argument specifically in relation to the school achievement of Black students. They argued that once "rumors of inferiority" (stereotypes; p. 18) about Black students' abilities pervade the environment—through, for example, national debates over the genetic basis of racial differences in IQ—they can intimidate Black students; become internalized by them; and, in turn, lead to a low sense of self-efficacy, demotivation, and underperformance in school. Analogous arguments have been applied to women interested in math-related areas (cf. Eccles-Parsons et al., 1983).

These models recognize the situational influence of negative stereotypes (e.g., Allport, 1954; Howard & Hammond, 1985; S. Steele, 1990) but most often describe it as a process in which the stereotype, or more precisely the possibility of being stereotyped, triggers an internalized inferiority doubt or low expectancy. And because this anxiety is born of a socialization presumed to influence all members of the stereotyped group, virtually all members of the group are presumed to have this anxiety, to one degree or another.

Stereotype threat, in contrast, refers to the strictly situational threat of negative stereotypes, the threat that does not depend on cuing an internalized anxiety or expectancy. It is cued by the mere recognition that a negative group stereotype could apply to oneself in a given situation. How threatening this recognition becomes depends on the person's identification with the stereotype-relevant domain. For the domain identified, the situational relevance of the stereotype is threatening because it threatens diminishment in a domain that is self-definitional. For the less domain identified, this recognition is less threatening or not threatening at all, because it threatens something that is less self-definitional.

Stereotype threat, then, as a situational pressure "in the air" so to speak, affects only a subportion of the stereotyped group and, in the area of schooling, probably affects confident students more than unconfident ones. Recall that to be identified with schooling in general, or math in particular, one must have confidence in one's domain-related abilities, enough to perceive good prospects in the domain. This means that stereotype threat should have its greatest effect on the better, more confident students in stereotyped groups, those who have not internalized the group stereotype to the point of doubting their own ability and have thus remained identified with the domain—those who are in the academic vanguard of their group.[2]

Several general features of stereotype threat follow:

1. Stereotype threat is a general threat not tied to the psychology of particular stigmatized groups. It affects the members of any group about whom there exists some generally known negative stereotype (e.g., a grandfather who fears that any faltering of memory will confirm or expose him to stereotypes about the aged). Stereotype threat can be thought of as a subtype of the threat posed by negative reputations in general.

2. That which turns stereotype threat on and off, the controlling "mechanism" so to speak, is a particular concurrence: whether a negative stereotype about one's group becomes relevant to interpreting oneself or one's behavior in an identified-with setting. When such a set-

[2] The point is not that negative stereotypes are never internalized as low self-expectancies and self-doubts. It is that in such internalization, disidentification is the more primary adaptation. That is, once the stereotype-relevant domain (e.g., math) is dropped as a self-definition, the negative stereotype (e.g., that women are limited in math) can be accepted as more self-descriptive (i.e., internalized) without it much affecting one's self-regard (as for the woman who, not caring about math, says she is lousy at it). But this internalization is probably resisted (e.g., Crocker & Major, 1989) until disidentification makes it less self-threatening. Once this has happened, the person is likely to avoid the domain because of both disinterest and low confidence regardless of whether stereotype threat is present.

ting integrates stereotyped and nonstereotyped people, it may make the stereotype, as a dimension of difference, more salient and thus more strongly felt (e.g., Frable, Blackstone, & Sherbaum, 1990; Goffman, 1963; Kleck & Strenta, 1980; Sartre, 1946/1965; S. Steele, 1990). But such integration is neither necessary nor sufficient for this threat to occur. It can occur even when the person is alone, as for a woman taking an important math test alone in a cubicle but under the threat of confirming a stereotyped limitation of ability. And, in integrated settings, it need not occur. Reducing the interpretive relevance of a stereotype in the setting, say in a classroom or on a standardized test, may reduce this threat and its detrimental effects even when the setting is integrated.[3]

3. This mechanism also explains the variabilities of stereotype threat: the fact that the type and degree of this threat vary from group to group and, for any group, across settings. For example, the type and degree of stereotype threat experienced by White men, Black people, and people who are overweight differ considerably, bearing on sensitivity and fairness in the first group, on school performance in the second, and on self-control in the third. Moreover, for any of these groups, this threat will vary across settings (e.g., Goffman, 1963; S. Steele, 1990). For example, women may reduce their stereotype threat substantially by moving across the hall from math to English class. The explanation of this model is straightforward: Different groups experience different forms and degrees of stereotype threat because the stereotypes about them differ in content, in scope, and in the situations to which they apply.

4. To experience stereotype threat, one need not believe the stereotype nor even be worried that it is true of oneself. The well-known African American social psychologist James M. Jones (1997) wrote,

> When I go to the ATM machine and a woman is making a transaction, I think about whether she will fear I may rob her. Since I have no such intention, how do I put her at ease? Maybe I can't . . . and maybe she has no such expectation. But it goes through my mind. (p. 262)

Jones felt stereotype threat in this situation even though he did not believe that the stereotype characterized him. Of course, this made it no less a life-shaping force. One's daily life can be filled with recurrent situations in which this threat pressures adaptive responses.

5. The effort to overcome stereotype threat by disproving the stereotype—for example, by outperforming it in the case of academic work—can be daunting. Because these stereotypes are widely disseminated throughout society, a personal exemption from them earned in one setting does not generalize to a new setting where either one's reputation is not known or where it has to be renegotiated against a new challenge. Thus, even when the stereotype can be disproven, the need to do so can seem Sisyphean, everlastingly recurrent. And in some critical situations, it may not be disprovable. The stereotypes considered in this work allege group-based limitations of ability that are often reinforced by the structural reality of increasingly small group representations at more advanced levels of the schooling domain. Thus, for group members working at these advanced levels, no amount of success up to that point can disprove the stereotype's relevance to their next, more advanced performance. For the advanced female math student who has been brilliant up to that point, any frustration she has at the frontier of her skills could confirm the gender-based limitation alleged in the stereotype, making this frontier, because she is so invested in it, a more threatening place than it is for the nonstereotyped. Thus, the work of dispelling stereotype threat through performance probably increases with the difficulty of work in the domain, and whatever exemption is gained has to be rewon at the next new proving ground.

Empirical Support for a Theory of Stereotype Threat and Disidentification

In testing these ideas, the research of my colleagues and I has had two foci: The first is on intellectual performance in the domain in which negative group stereotypes apply. Here, the analysis has two testable implications. One is that for domain-identified students, stereotype threat may interfere with their domain-related intellectual performance. Analysts have long argued that behaving in a situation in which one is at risk of confirming a negative stereotype about one's group, or of being seen or treated stereotypically, causes emotional distress and pressure (e.g., Cross, 1991; Fanon, 1952/1967; Goffman, 1963; Howard & Hammond, 1985; Sartre, 1946/1965; C. M. Steele & Aronson, 1995; S. Steele, 1990). The argument here is that for those who identify with the domain enough to experience this threat, the pressure it causes may undermine their domain performance. Disruptive pressures such as evaluation apprehension, test anxiety, choking, and token status have long been shown to disrupt performance through a variety of mediating mechanisms: interfering anxiety, reticence to respond, distracting thoughts, self-consciousness, and so on (Baumeister & Showers, 1984; Geen, 1991; Lord & Saenz, 1985; Sarason, 1980; Wine, 1971). The assumption of this model is that stereotype threat is another such interfering pressure. The other testable implication is that reducing this threat in the performance setting, by reducing its interfering pressure, should improve the performance of otherwise stereotype-threatened students.

The second research focus is the model's implica-

[3] As a process of social devaluation, stereotype threat is both a subform of stigmatization and something more general. It is that form of stigmatization that is mediated by collectively held, devaluing group stereotypes. This means that it does not include stigmatization that derives from nonstereotyped features such as a facial disfigurement or, for example, what Goffman (1963) called abominations of the body. Stereotype threat is a situational predicament. And, in this sense, it is also more general than stigmatization. It is a threat that can befall anyone about whom a negative reputation or group stereotype exists.

tion that stereotype threat, and the anticipation of having to contend with it unceasingly in school or some domain of schooling, should deter members of these groups from identifying with these domains, and, for group members already identified, it should pressure their disidentification.[4]

Stereotype Threat and Intellectual Performance

Steven Spencer, Diane Quinn, and I (Spencer et al., 1997) first tested the effect of stereotype threat on intellectual performance by testing its effect on the standardized math test performance of women who were strong in math.

The stereotype threat of women performing math. At base, of course, the stereotype threat that women experience in math-performance settings derives from a negative stereotype about their math ability that is disseminated throughout society. But whether this threat impaired their performance, we reasoned, would depend on two things. First, the performance would have to be construed so that any faltering would imply the limitation of ability alleged in the stereotype. This means that the performance would have to be difficult enough so that faltering at it would imply having reached an ability limit but not so difficult as to be nondiagnostic of ability. And second, as has been much emphasized, the women in question would have to be identified with math, so that faltering and its stereotype-confirming implication would threaten something they care about, their belongingness and acceptance in a domain they identify with. Of course, men too (at least those of equal skill and identification with math) could be threatened in this situation; faltering would reflect on their ability too. But their faltering would not carry the extra threat of confirming a stereotyped limitation in math ability or of causing them to be seen that way. Thus, the threat that women experience, through the interfering pressure it causes, should worsen their performance in comparison to equally qualified men. Interestingly, though, these otherwise confident women should perform equally as well as equally qualified men when this situational threat is lessened.

To explore these questions, Spencer, Quinn, and I (Spencer et al., 1997) designed a basic research paradigm: We recruited female and male students, mostly college sophomores, who were both good at math and strongly identified with it in the sense of seeing themselves as strong math students and seeing math as important to their self-definition. We then gave them a very difficult math test one at a time. The items were taken from the advanced math General Records Examination (GRE) and we assumed would frustrate the skills of these students without totally exceeding them. As expected, and presumably reflecting the impairing effects of stereotype threat, women significantly underperformed in relation to equally qualified men on this difficult math test. But more important, in another condition of this experiment in which the test was an advanced literature test rather than a math test and in which participants had been selected and matched for their strong literature skills and identification, women performed just as well as equally qualified men. This happened, we reasoned, because women are not stereotype threatened in this area.

A second experiment replicated women's underperformance on the difficult math test and showed that it did not happen when the test was easier, that is when the items, taken from the regular quantitative section of the GRE, were more within the skills of these strong math students. The lack of performance frustration on this easier test, presumably, reduced women's stereotype threat by making the stereotype less relevant as an interpretation of their performance.

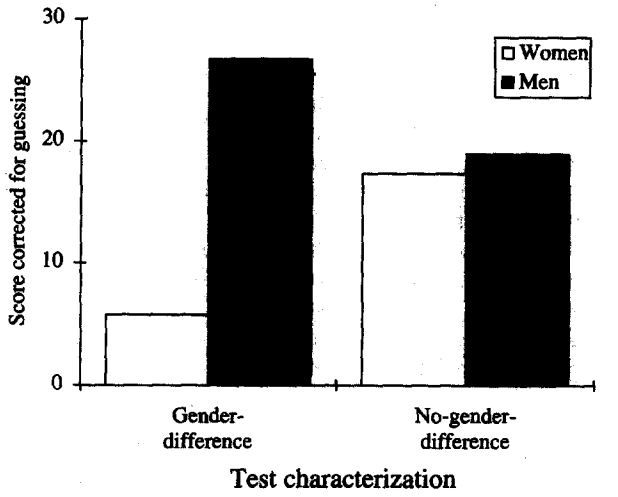
Stereotype threat versus genes. So went our interpretation. But an alternative was possible: The biological limits of women's math ability do not emerge until the material tested is difficult. It is this very pattern of evidence that Benbow and Stanley (1980, 1983) used to suggest a genetic limitation in women's math ability. Thus, the first two experiments reproduced the gender effects on math performance reported in the literature: that women underperform primarily in math and mainly when the material is difficult. But they fall short of establishing our interpretation.

To do this, we would need to give women and men a difficult math test (one capable of producing women's underperformance) but then experimentally vary stereotype threat, that is, vary how much women were at risk of confirming the stereotype while taking the test. A third experiment did this by varying how the test (the same difficult one used in the earlier experiments) was represented. Participants were told either that the test generally showed gender differences, implying that the stereotype of women's limitations in math was relevant to interpreting their own frustration, or that it showed no gender differences, implying that the gender stereotype was not relevant to their performance and thus could not be confirmed by it on this particular test. The no-gender-differences representation did not challenge the validity of the stereotype; it simply eliminated the risk that the stereotype could be fulfilled on this test. In the gender-differences condition, we expected women (still stereotype threatened) to underperform in relation to equally qualified men, but in the no-gender-differences condition, we expected women (with stereotype threat reduced) to perform equal to such men. The genetic interpretation, of course, predicts that women will underperform on this difficult test regardless of how it is represented.

In dramatic support of our reasoning, women performed worse than men when they were told that the test produced gender differences, which replicated women's

[4] Moreover, a protective avoidance of identification can become a group norm. In reaction to a shared sense of threat in school, for example, it can become a shared reaction that is transmitted to group members as the normative relation to school. Both research (e.g., Ogbu, 1986; Solomon, 1992) and the media have documented this reaction in minority students from inner-city high schools to Harvard University's campus. Thus, disidentification can be sustained by normative pressure from the in-group as well as by stereotype threat in the setting.

Figure 1
Mean Performance on a Difficult Math Test as a Function of Gender and Test Characterization



underperformance observed in the earlier experiments, but they performed equal to men when the test was represented as insensitive to gender differences, even though, of course, the same difficult "ability" test was used in both conditions (see Figure 1). Genetic limitation did not cap the performance of women in these experiments. A fourth experiment showed that reducing stereotype threat (through the no-gender-differences treatment) raised women's performance to that of equally qualified men, even when participants' specific performance expectancies were set low, that is, when participants were led to expect poor test performance. Also, a fifth experiment (that again replicated the treatment effects of the third experiment) found that participants' posttreatment anxiety, not their expectancies or efficacy, predicted their performance. Thus, the disruptive effect of stereotype threat was mediated more by the self-evaluative anxiety it caused than by its lowering of performance expectations or self-efficacy.

Internal or situational threat. These findings make an important theoretical and practical point: The gender-differences conditions (including those in which the possibility of gender differences was left to inference rather than stated directly) did not impair women's performance by triggering doubts they had about their math ability. For one thing, these women had no special doubts of this sort; they were selected for being very good at math and for reporting high confidence in their ability. Nor was this doubt a factor in their test performance. Recall that the math test was represented as an ability test in all conditions of these experiments. This means that in the no-gender-differences conditions, women were still at risk of showing their own math ability to be weak—the same risk that men had in these conditions. Under this risk (when their own math ability was on the line), they performed just as well as men. Whatever performance-impairing anxiety they had, it was no greater than that of equally qualified men. Thus, the gender-differences conditions (the normal condition under which people take these tests) could not have impaired their performance by triggering some greater internalized anxiety that women have about their own math ability—an anxiety acquired, for example, through prior socialization. Rather, this condition had its effect through situational pressure. It set up an interpretive frame such that any performance frustration signaled the possible gender-based ability limitation alleged in the stereotype. For these women, this signal challenged their belongingness in a domain they cared about and, as a possibly newly met limit to their ability, could not be disproven by their prior achievements, thus its interfering threat.

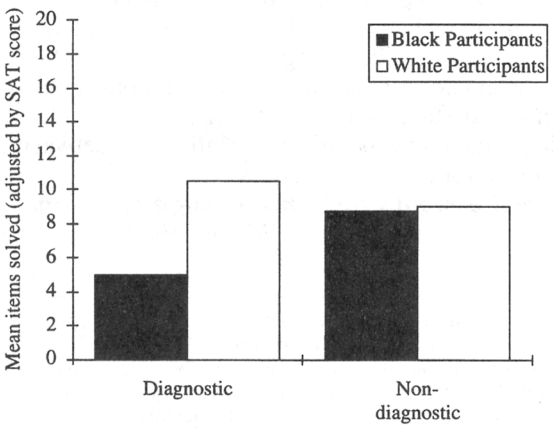
The stereotype threat of African Americans on standardized tests. Joshua Aronson and I (C. M. Steele & Aronson, 1995) examined these processes among African American students. In these studies, Black and White Stanford University students took a test composed of the most difficult items on the verbal GRE exam. Because the participants were students admitted to a highly selective university, we assumed that they were identified with the verbal skills represented on standardized tests. The first study varied whether or not the stereotype about Black persons' intellectual ability was relevant to their performance by varying whether the test was presented as *ability-diagnostic,* that is, as a test of intellectual ability, or as *ability-nondiagnostic,* that is, as a laboratory problem-solving task unrelated to ability and thus to the stereotype about ability. Analysis of covariance was used to remove the influence of participants' initial skills, measured by their verbal SAT scores, on their test performance. This done, the results showed strong evidence of stereotype threat: Black participants greatly underperformed White participants in the diagnostic condition but equaled them in the nondiagnostic condition (see Figure 2). A second experiment produced the same pattern of results with an even more slight manipulation of stereotype threat: whether or not participants recorded their race on a demographic questionnaire just before taking the test (described as nondiagnostic in all conditions). Salience of the racial stereotype alone was enough to depress the performance of identified Black students (see Figure 3).

The cognitive mediation of stereotype threat. Stereotype threat, then, can impair the standardized test performance of domain-identified students; this effect generalizes to several ability-stereotyped groups, and its mediation seems to involve anxiety more than expectancies. But do these manipulations cause a specific state of stereotype threat, that is, a sensed threat specifically about being stereotyped or fitting the stereotype? To address this question, Aronson and I (C. M. Steele & Aronson, 1995) tested two things: whether manipulating stereotype threat actually activates the racial stereotype in

the thinking and information processing of stereotype-threatened test takers and whether it produces in them a specific motivation to avoid being seen stereotypically. Again, Black and White participants were run in either an ability-diagnostic or ability-nondiagnostic condition, except that just after the condition instructions and completion of the sample test items (so that participants could see how difficult the items were) and just before participants expected to take the test, they completed measures of stereotype activation and avoidance. The stereotype-activation measure asked them to complete 80 word fragments, 10 of which we knew from pretesting could be completed with, among other words, words symbolic of African American stereotypes (e.g., _ _ce [race], la_ _ [lazy], or _ _or [poor]) and 5 of which could be completed with, among other words, words signifying self-doubts (e.g., lo_ _ _ [loser], du_ _ [dumb], or sha_ _ _ [shame]). The measure of participants' motivation to avoid being seen stereotypically simply asked them how much they preferred various types of music, activities, sports, and personality traits, some of which a pretest sample had rated as stereotypic of African Americans.[5]

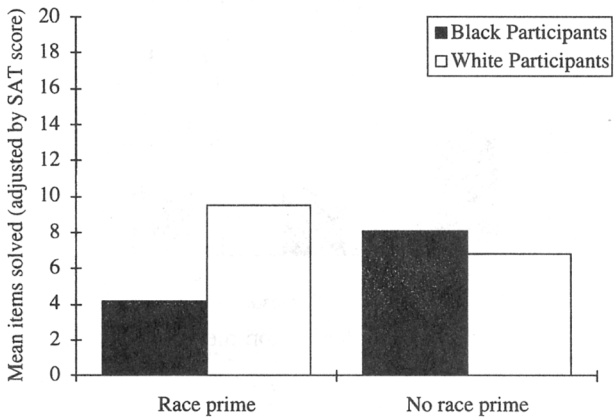
If expecting to take a difficult ability-diagnostic test is enough to activate the racial stereotype in the thinking of Black participants and to motivate them to avoid being stereotyped, then these participants, more than those in the other conditions, should show more stereotype and self-doubt word completions and fewer preferences for things that are African American. This is precisely what happened. Black participants in the diagnostic condition completed more word fragments with stereotype- and self-doubt-related words and had fewer preferences for things related to African American experience (e.g., jazz, basketball, hip-hop) than Black participants in the nondiagnostic condition or White participants in either condition, all of whom were essentially the same (see Figure 4). Also, as a last item before participants expected to begin the test, they were given the option of recording their race, a measure we thought might further tap into an apprehension about being viewed stereotypically. Interestingly, then, all of the Black participants in the nondiagnostic condition and all of the White participants in both conditions listed their race, whereas only 25% of the Black participants in the diagnostic condition did so.

Figure 2
Mean Performance on a Difficult Verbal Test as a Function of Race and Test Characterization



Note. SAT = Scholastic Assessment Test.

Figure 3
Mean Performance on a Difficult Verbal Test as a Function of Whether Race Was Primed

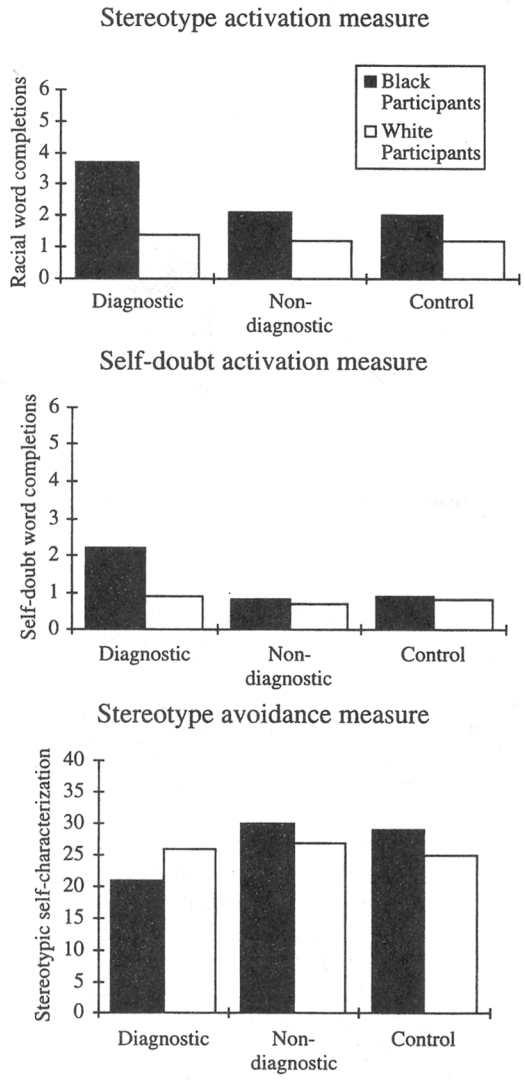


Note. SAT = Scholastic Assessment Test.

Self-rejection or self-presentation? A troubling implication of the earlier mentioned internalization models (e.g., Allport, 1954; Bettelheim, 1943; Clark, 1965; Grier & Coobs, 1968; Erikson, 1956; Fanon, 1952/1967; Kardiner & Ovesey, 1951) is that negative stereotypes about one's group eventually become internalized and cause rejection of one's own group, even of oneself—*self-hating* preferences. The famous finding of Clark and Clark (1939) that Black children preferred White dolls over Black dolls has been interpreted this way. The preferences of Black participants in the diagnostic condition fit this pattern; with negative stereotypes about their group cognitively activated, they valued things that were African American less than any other group. But the full set of results suggests a different interpretation. In those conditions in which Black participants did not have to worry about tripping a stereotypic perception of themselves, they valued things that were African American more strongly than did other participants. Thus, rather than reflecting self- or own-group rejection, their devaluing of things that were African American in the diagnostic

[5] Participants did not actually take the test in this experiment, as completing these measures would likely have activated the stereotype in all conditions.

Figure 4
Indicators of Stereotype Threat



condition was apparently a strategic self-presentation aimed at cracking the stereotypic lens through which they could be seen. So it could be, then, in the general case, rather than reflecting real self-concepts, behavior that appears group rejecting or self-rejecting may reflect situation-bound, self-presentational strategies.

Stereotype threat and domain identification. Not being identified with a domain, our (C. M. Steele & Aronson, 1995) theory reasons, means that one's experience of stereotype threat in the domain is less self-threatening. Although we have yet to complete a satisfactory test of this prediction, partially completed experiments and pretests show that stereotype threat has very little, if any, effect on participants not identified with the domain of relevance. Most typically, these participants give up and underperform on the difficult test regardless of whether they are under stereotype threat. Although not yet constituting a complete test of this implication of the theory, these data do emphasize that the above results generalize only to domain-identified students.

Stereotype threat and the interpretation of group differences in standardized test performance. Inherent to the science of quantifying human intelligence is the unsavory possibility of ranking societal groups as to their aggregated intelligence. It is from this corner of psychology that the greatest controversy has arisen, a controversy that has lasted throughout this century and that is less about the fact of these group differences than about their interpretation (cf. Herrnstein & Murray, 1994; Kamin, 1974). To the set of possible causes for these group differences, our (C. M. Steele & Aronson, 1995) findings add a new one: the differential impact of stereotype threat on groups in the testing situation itself. Thus, stereotype threat may be a possible source of bias in standardized tests, a bias that arises not from item content but from group differences in the threat that societal stereotypes attach to test performance. Of course, not every member of an ability-stereotyped group is going to be affected by stereotype threat every time they take a test. As our research has shown, the experience of success as one takes the test can dispel the relevance of the stereotype. Nonetheless, among the most identified test takers in the stereotype-threatened group—those in its academic vanguard who have the greatest confidence and skills—this threat can substantially depress performance on more difficult parts of the exam. And this depression could contribute significantly to the group's underperformance in comparison with nonstereotype-threatened groups.[6]

Reaction of Disidentification

Stereotype threat is assumed to have an abiding effect on school achievement—an effect beyond its impairment of immediate performance—by preventing or breaking a person's identification with school, in particular, those domains of schooling in which the stereotype applies. This reasoning has several implications for which empirical evidence can be brought to bear: the resilience of self-esteem to stigmatization; the relationship between stigmatized status and school achievement; and, among ability-stigmatized people, the relationship between their school performance and self-esteem.

Self-esteem's resilience to stigmatization. In a recent review, Crocker and Major (1989) were able to

[6] Those who are less domain identified in the stereotype-threatened group may also underperform on standardized tests. Because they care less about the domain it represents, they may be undermotivated or they may withdraw effort in the face of frustration. And for all of the reasons I have discussed, the greater portion of the stereotype-threatened group may be academically unidentified. This fact too, then, may contribute to the group's overall weaker performance on these tests in comparison with nonstereotype-threatened groups.

make a strong case for the lack of something that common sense suggests should exist: a negative effect of stigmatization on self-esteem. Following the logic of the internalization models described above and viewing stigmatization as, among other things, an assault to self-esteem, one might expect that people who are stigmatized would have lower self-esteem than people who are not. Yet, as Crocker and Major reported, when the self-esteem of stigmatized groups (e.g., Blacks, Chicanos, the facially disfigured, obese people, etc.) is actually measured, one finds that their self-esteem is as high as that of the nonstigmatized.

Crocker and Major (1989) offered the intriguing argument that stigma itself offers esteem-protective strategies. For example, the stigmatized can blame their failures on the prejudice of out-group members, they can limit their self-evaluative social comparisons to the in-group of other stigmatized people, and they can devalue the domains in which they feel devalued. Other models have also described esteem-saving adaptations to stigma. For example, models that assume internalization of stereotype-related anxieties often posit compensatory personality traits (e.g., grandiosity) that protect self-esteem but leave one poorly adapted to the mainstream (e.g., Allport, 1954; Clark, 1965; Grier & Coobs, 1968; Kardiner & Ovesey, 1951; S. Steele, 1990). In the present reasoning, stigmatization stems from stereotype threat in specific domains. Thus, it adds to the list of stigma adaptations the possibility of simple domain disidentification, the rescuing of self-esteem by rendering as self-evaluatively irrelevant the domain in which the stereotype applies. Herein may lie a significant source of the self-esteem resilience shown in stigmatized groups. This idea also implies that once domain disidentification is achieved, the pressure for adaptations of attribution and personality may be reduced.

A universal connection between stigmatization and poor school achievement. If disidentification with school, and the resulting underachievement, can be a reaction to ability-stigmatizing stereotypes in society, then it might be expected that ability stigmatization would be associated with poor school performance wherever it occurs in the world. Finding such a relationship would not definitively prove the present theory; the direction of causality could be quarreled with, as could the mediation of such a relationship. Still, it would be suggestive, and, in that respect, Ogbu (1986) reported an interesting fact: Among the caste-like minorities in industrial and nonindustrial nations throughout the world (e.g., the Maoris of New Zealand, the Baraku of Japan, the Harijans of India, the Oriental Jews of Israel, and the West Indians of Great Britain), there exists the same 15-point IQ gap between them and the nonstigmatized members of their society as exists between Black and White Americans. These groups also suffer poorer school performance, higher dropout rates, and related behavior problems. Moreover, these gaps appear even when the stigmatized and nonstigmatized are of the same race, as in the case of the Baraku and other Japanese. What these groups share that is capable of explaining their deficits is a caste-like status that, through stereotypes in their societies, stigmatizes their intellectual abilities—sowing the seeds, I suggest, of their school disidentification.

The disassociation of self-esteem and school achievement. If the poor school achievement of ability-stigmatized groups is mediated by disidentification, then it might be expected that among the ability stigmatized, there would be a disassociation between school outcomes and overall self-esteem. Several kinds of evidence suggest this process among African Americans. First, there is the persistent finding that although Black students underperform in relation to White students on school outcomes from grades to standardized tests (e.g., Demo & Parker, 1987; Simmons, Brown, Bush, & Blyth, 1978; C. M. Steele, 1992), their global self-esteem is as high or higher than that of White students (e.g., Porter & Washington, 1979; Rosenberg, 1979; Wylie, 1979). For both of these facts to be true, some portion of Black students must have acquired an imperviousness to poor school performance.

Several further studies suggest that this imperviousness is rooted in disidentification. In a study of desegregated schools in Champaign, Illinois, Hare and Costenell (1985) measured students' school achievement; overall self-esteem; and self-esteem in the specific domains of home life, school, and peer-group relations. Like others, they found that although Black students performed less well than White students, they still had comparable levels of overall self-esteem. Their domain-specific measures suggested why: Although Black students were lower than White students in school and home-life self-esteem, Blacks slightly exceeded Whites in peer-group self-esteem. Here then, perhaps, was the source of their overall self-regard: disidentification with domains in which their evaluative prospects were poor (in this case, school and home life) and identification with domains in which their prospects were better (i.e., their peers).

A recent study suggests that this may be a not uncommon phenomenon. Analyzing data available from the National Educational Longitudinal Survey (National Center for Educational Statistics, 1992; a nationally representative longitudinal survey begun in 1988), Osborne (1994) found that from the 8th through 10th grades, Black students had lower achievement and somewhat higher self-esteem than White students, which replicated the general pattern of findings described above. But more than this, he found evidence of increasing Black students' disidentification over this period: The correlation between their school achievement and self-esteem for this period decreased significantly more for Black than for White students. Also, using a scale measure of school disidentification, Major, Spencer, Schmader, Wolfe, and Crocker (in press) found that Black students were more disidentified than White students in several college samples and that for disidentified students of both races, negative feedback about an intellectual task had less effect on their self-esteem than it did for identified students.

Major et al. further showed that when racial stereotypes were primed, neither negative nor positive feedback affected Black students' self-esteem, whereas the self-esteem of White students followed the direction of the feedback. Ability stigmatization of the sort experienced by African Americans, then, can be associated with a protective "disconnect" between performance and self-regard, a disconnect of the sort that is consistent with disidentification theory.

Can stereotype threat directly cause this disconnect? To test this question, Kirsten Stoutemeyer and I varied the strength of stereotype threat that female test takers (Stanford students) were under by varying whether societal differences between women and men in math performance were attributed to small but stable differences in innate ability (suggesting an inherent, gender-based limit in math ability) or to social causes such as sex-role prescriptions and discrimination (suggesting no inherent, gender-based limit in math ability). We then measured their identificaion with math and math-related careers, either before or after they took a difficult math test. Regardless of when identification was measured, women under stronger stereotype threat disidentified with math and math-related careers more than women under weaker stereotype threat. Although domain identification has several determinants, these findings suggest that stereotype threat is an important one of them.

"Wise" Schooling: Practice and Policy

As a different diagnosis, the present analysis comes to a different prescription: The schooling of stereotype-threatened groups may be improved through situational changes (analogous to those manipulated in our experiments) that reduce the stereotype threat these students might otherwise be under. As noted, psychological diagnoses have more typically ascribed the problems of these students to internal processes ranging from genes to internalized stereotypes. On the face of it, at least, internal states are more difficult to modify than situational factors. Thus, the hope of the present analysis, encouraged by our research, is that these problems might be more tractable through the situational design of schooling, in particular, design that secures these students in the belief that they will not be held under the suspicion of negative stereotypes about their group. Schooling that does this, I have called *wise,* a term borrowed from Irving Goffman (1963), who borrowed it from gay men and lesbians of the 1950s. They used it to designate heterosexuals who understood their full humanity despite the stigma attached to their sexual orientation: family and friends, usually, who knew the person beneath the stigma. So it must be, I argue, for the effective schooling of stereotype-threatened groups.

Although "wisedom" may be necessary for the effective schooling of such students, it may not always be sufficient. The chief distinction made in this analysis (between those of these groups who are identified with the relevant school domain and those who are not) raises a caution. As noted, stereotype threat is not keenly felt by those who identify little with the stereotype-threatening domain. Thus, although reducing this threat in the domain may be necessary to encourage their identification, it may not be sufficient to build an identification that is not there. For this to occur, more far-reaching strategies that develop the building blocks of domain identification may be required: better skills, greater domain self-efficacy, feelings of social and cultural comfort in the domain, a lack of social pressure to disidentify, and so on.

But for the identified of these groups, who are quite numerous on college campuses, the news may be better than is typically appreciated. For these students, feasible changes in the conditions of schooling that make threatening stereotypes less applicable to their behavior (i.e., wisdom) may be enough. They are already identified with the relevant domain, they have skills and confidence in the domain, and they have survived other barriers to identification. Their remaining problem is stereotype threat. Reducing that problem, then, may be enough to bring their performance on par with that of nonstereotyped persons in the domain.

This distinction raises an important and often overlooked issue in the design of schooling for stereotype-threatened students, that of *triage,* the issue of rendering onto the right students the right intervention. Mistakes can easily be made. For example, applying a strategy to school-identified students (on the basis of their membership in a stereotype-threatened group) that assumes weak identification, poor skills, and little confidence could backfire. It could increase stereotype threat and underperformance by signaling that their abilities are held under suspicion because of their group membership. But the opposite mistake could be made by applying a strategy that assumes strong identification, skills, and confidence to those who are actually unidentified with the relevant domain. Merely reducing stereotype threat may not accomplish much when the more primary need of these students is to gain the interests, resources, skills, confidences, and values that are needed to identify with the domain.

Some wise strategies, then, may work for both identified and unidentified students from these groups, but others may have to be appropriately targeted to be effective. I offer some examples of both types.

For both domain-identified and domain-unidentified students:

1. Optimistic teacher–student relationships. The prevailing stereotypes make it plausible for ability-stigmatized students to worry that people in their schooling environment will doubt their abilities. Thus, one wise strategy, seemingly suitable for all students, is to discredit this assumption through the authority of potential-affirming adult relationships. The Comer (1988) Schools Project has used this strategy with great success at the elementary school level, and Johnides, von Hippel, Lerner, and Nagda (1992) have used it in designing a mentoring program for incoming minority and other students

at the University of Michigan. In analogous laboratory experiments, Geoffrey Cohen, Lee Ross, and I (Cohen, Steele, & Ross, 1997) found that critical feedback to African American students was strongly motivating when it was coupled with optimism about their potential.

2. Challenge over remediation. Giving challenging work to students conveys respect for their potential and thus shows them that they are not regarded through the lens of an ability-demeaning stereotype. Urie Treisman (1985) used this strategy explicitly in designing his successful group-study workshops in math for college-aged women and minorities. Taking students where they are skillwise, all students can be given challenging work at a challenging, not overwhelming, pace, especially in the context of supportive adult–student relationships. In contrast, remedial work reinforces in these students the possibility that they are being viewed stereotypically. And this, by increasing stereotype threat in the domain, can undermine their performance.

3. Stressing the expandability of intelligence. The threat of negative-ability stereotypes is that one could confirm or be seen as having a fixed limitation inherent to one's group. To the extent that schooling can stress what Carol Dweck (1986) called the *incremental* nature of human intelligence—its expandability in response to experience and training—it should help to deflect this meanest implication of the stereotype. Aronson (1996) recently found, for example, that having African American college students repeatedly advocate the expandability of intelligence to their elementary school tutees significantly improved their own grades.

For domain-identified students:

1. Affirming domain belongingness. Negative-ability stereotypes raise the threat that one does not belong in the domain. They cast doubt on the extent of one's abilities, on how well one will be accepted, on one's social compatibility with the domain, and so on. Thus, for students whose primary barrier to school identification is stereotype threat, direct affirmation of their belongingness in the domain may be effective. But it is important to base this affirmation on the students' intellectual potential. Affirming social belonging alone, for those under the threat of an ability stereotype, could be taken as begging the question.

2. Valuing multiple perspectives. This refers to strategies that explicitly value a variety of approaches to both academic substance and the larger academic culture in which that substance is considered. Making such a value public tells stereotype-threatened students that this is an environment in which the stereotype is less likely to be used.

3. Role models. People from the stereotype-threatened group who have been successful in the domain carry the message that stereotype threat is not an insurmountable barrier there.

For domain-unidentified students:

1. Nonjudgmental responsiveness. Research by Lepper, Woolverton, Mumme, and Gurtner (1993) has identified a distinct strategy that expert tutors use with especially poor students: little direct praise, Socratic direction of students' work, and minimal attention to right and wrong answers. For students weakly identified with the domain, who are threatened by a poor reputation and who probably hold internalized doubts about their ability, this Socratic strategy has the wisedom of securing a safe teacher–student relationship in which there is little cost of failure and the gradual building of domain efficacy from small gains.

2. Building self-efficacy. Based on Bandura's (1977, 1986) theory of self-efficacy, this strategy attempts to build the student's sense of competence and self-efficacy in the schooling domain. Howard and Hammond (1985) have developed a powerful implementation of this strategy for African American and other minority students, especially in inner-city public schools.

Existence Proof: A Wise Schooling Intervention

Providing a definitive test of wise schooling theory will require, of course, an extensive research program. But as a first step, something might be learned from what Urie Treisman (1985) called an existence proof, in this case, a demonstration that an intervention derived from the theory could stop or reverse a tenacious negative trajectory in the school performance of stereotype-threatened students. Such an intervention would of necessity confound things: different wise practices as well as other practices and structures, peculiar to that setting, that could also affect academic outcomes. It could not stand as a test of the psychological theory per se. But if a particular architecture of wise strategies succeeded, it would encourage their applicability to the real-world schooling of these students.

With this rationale, my colleagues and I (Steven Spencer, Richard Nisbett, Mary Hummel, David Schoem, Kent Harber, Ken Carter) implemented a freshman-year program at the University of Michigan aimed at the underachievement and low retention rates of African American students. Each year, the program included approximately 250 freshmen in the ethnic proportions of the larger campus but with an oversampling of approximately 20% Black students and 20% non-Black minority students (i.e., Asian, Hispanic, and Native American students as a single group). Program students were randomly selected from the students admitted to Michigan and then recruited by phone to participate. All program participants lived together in the wing of a large, 1,200-student dormitory throughout their freshman year.

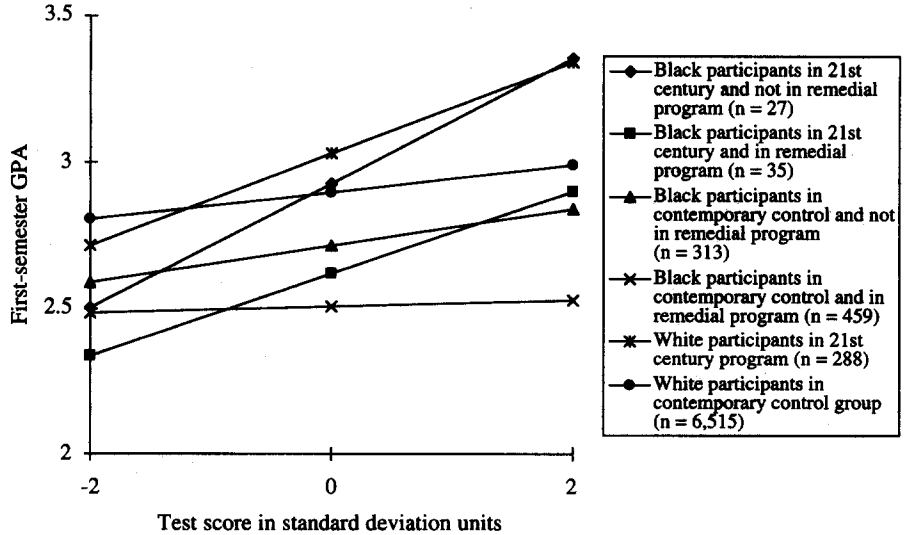
In this context, we implemented several wise strategies. The program was presented as a transition program aimed at helping students maximize the advantages of university life. We also recruited students honorifically; they were told that, as Michigan admittees, they had survived a very competitive selection process and that our program was designed to help them maximize their strong potential. These practices represented the program

as nonremediational and represented the university as having acknowledged their intellectual potential and as having high expectations for them—all things that signal the irrelevance of negative group stereotypes. Once the students were in the program, these expectations were reinforced by their being offered a "challenge" workshop, modeled on those developed by Treisman (1985) for calculus, in either freshman calculus, chemistry, physics, or writing. These were taken on a voluntary basis in the dormitory. Students also participated in small weekly discussion groups, centered on brief readings, that allowed discussion of adjustment-relevant social and even personal issues. This activity has the wisdom of letting students know that they, or other members of their group, are not the only ones with concerns about adjusting to university life—an insight that can deflect the relevance of negative group stereotypes. These formal program components lasted for the first 10 weeks of the school year, and, as voluntary activities, approximately half of the students regularly participated in either one or both of them.

The first-semester grades averaged over the first two years of this ongoing project give a reliable picture of the program's initial impact. To show the size of the program's effect on students at different levels of preparation, Figure 5 graphs first-semester grades, using regression lines, for the different student groups as a function of standardized test scores on entry into the university (they are presented as standard deviation units in this figure to provide a common scale for students who took either the SAT or American College Test exam). The first thing to notice is the two essentially parallel lines for White and Black students outside of any program at Michigan. They replicate the standard overprediction–underperformance of Black students alluded to earlier, and it is against this pattern that the effects of the program can be evaluated. Looking first at the line for White students in our program, there is a modest tendency for these students to do better than the White control students (i.e., those outside the program), but given our accumulation of n throughout these first two years, this difference is not significant. It is the results for Black students in our program (but who were not also in the campus minority program) that are most promising. Their line is considerably above that for Black control students (i.e., Black students outside any program) and, even with the modest sample size ($n = 27$), is significantly higher than this control line in the top one third of the standardized test distribution, $t = 2.72$, $p < .05$. It is important that this group of Black students showed almost no underperformance; in the top two thirds of the test distribution, they had essentially the same grades as White students. We also know from follow-up data that their higher grade performance continued at least through their sophomore year and that as long as four years later, only one of them had dropped out.

Theoretically just as important, is the bottom line in Figure 5, depicting the results for Black students in a large minority remediation program. Despite getting considerable attention, they performed worse than the other groups at nearly every level of preparation. The difference between Black students in the minority program and Black students not in any program becomes significant at 1.76 standard deviations below the mean for test performance and is significant from that point on, $ps < .05$. Also, by the beginning of their junior year,

Figure 5
First-Semester Grade Point Average (GPA) as a Function of Program and Race Controlling for High School GPA



25% of these students had failed to register, and among those who entered with test scores in the top one third of the test distribution, this figure was 40%. Some selection factor possibly contributed to this. Despite our having controlled for test scores and high school grade point averages in these analyses, some portion of these students may have been assigned to this program because they evidenced other risk factors. Still, these results suggest that the good intentions of the minority–remediation framework for schooling African American students can backfire by, in our terms, institutionalizing the racial stereotype by which they are already threatened.

Although these findings are preliminary and we do not know that they were mediated as our theory claims, they are a step toward an existence proof; they show that wise practices can reduce Black students' underachievement in a real-school context and, as important, that unwise practices seem to worsen it.

Conclusion

In social psychology, we know that as observers looking at a person or group, we tend to stress internal, dispositional causes of their behavior, whereas when we take the perspective of the actor, now facing the circumstances they face, we stress more situational causes (e.g., E. E. Jones & Nisbett, 1972; Ross, 1977). If there is a system to the present research, it is that of taking the actor's perspective in trying to understand the intellectual performance of African American and female students. It is this perspective that brings to light the broadly encompassing condition of having these groups' identification with domains of schooling threatened by societal stereotypes. This is a threat that in the short run can depress their intellectual performance and, over the long run, undermine the identity itself, a predicament of serious consequence. But it is a predicament—something in the interaction between a group's social identity and its social psychological context, rather than something essential to the group itself. Predicaments can be treated, intervened on, and it is in this respect that I hope the perspective taken in this analysis and the early evidence offer encouragement.

REFERENCES

Adelman, C. (1991). *Women at thirty-something: Paradoxes of attainment.* Washington, DC: U.S. Department of Education, Office of Research and Development.

Alexander, K. L., & Entwistle, D. R. (1988). Achievement in the first two years of school: Patterns and processes. *Monographs of the Society for Research in Child Development, 53*(2).

Allport, G. (1954). *The nature of prejudice.* New York: Doubleday.

American Council on Education. (1995–1996). *Minorities in higher education.* Washington, DC: Office of Minority Concerns.

Aronson, J. (1996). *Advocating the malleability of intelligence as an intervention to increase college grade performance.* Unpublished manuscript, University of Texas.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84,* 191–215.

Bandura, A. (1986). *Social foundations of action: A social–cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Baumeister, R. F., & Showers, C. J. (1984). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology, 16,* 361–383.

Benbow, C. P., & Arjmand, O. (1990). Predictions of high academic achievement in mathematics and science by mathematically talented students: A longitudinal study. *Journal of Educational Psychology, 82,* 430–441.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210,* 1262–1264.

Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222,* 1029–1031.

Bettelheim, B. (1943). Individual and mass behavior in extreme situations. *Journal of Abnormal and Social Psychology, 38,* 417–452.

Brill, A. A. (Ed.). (1938). *The basic writings of Sigmund Freud.* New York: Random House.

Burton, N. W., & Jones, L. V. (1982). Recent trends in achievement levels of Black and White youth. *Educational Researcher, 11,* 10–17.

Carter, S. (1991). *Reflections of an affirmative action baby.* New York: Basic Books.

Clark, K. B. (1965). *Dark ghetto: Dilemmas of social power.* New York: Harper & Row.

Clark, K. B., & Clark, M. K. (1939). The development of consciousness of self and the emergence of racial identification of Negro school children. *Journal of Social Psychology, 10,* 591–599.

Cohen, G., Steele, C. M., & Ross, L. (1997). *Giving feedback across the racial divide: Overcoming the effects of stereotypes.* Unpublished manuscript, Stanford University.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Comer, J. (1988, November). Educating poor minority children. *Scientific American, 259,* 42.

Cooley, C. H. (1956). *Human nature and the social order.* New York: Free Press.

Cose, E. (1993). *The rage of a privileged class.* New York: Harper Collins.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96,* 608–630.

Cross, W. E., Jr. (1991). *Shades of black: Diversity in African-American identity.* Philadelphia: Temple University Press.

Culotta, E., & Gibbons, A. (Eds.). (1992, November 13). Minorities in science [Special section]. *Science, 258,* 1176–1232.

DeBoer, G. (1984). A study of gender effects in science and mathematics course-taking behavior among students who graduated from college in the late 1970's. *Journal of Research in Science Teaching, 21,* 95–103.

Demo, D. H., & Parker, K. D. (1987). Academic achievement and self-esteem among Black and White college students. *Journal of Social Psychology, 4,* 345–355.

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41,* 1040–1048.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11,* 135–172.

Eccles-Parsons, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectations, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). New York: Freeman.

Erikson, E. (1956). The problem of ego-identity. *Journal of the American Analytical Association, 4,* 56–121.

Fanon, F. (1967). *Black skins, white masks.* New York: Grove Press. (Original work published 1952)

Frable, D., Blackstone, T., & Sherbaum, C. (1990). Marginal and mindful: Deviants in social interaction. *Journal of Personality and Social Behavior, 59,* 140–149.

Geen, R. G. (1991). Social motivation. *Annual Review of Psychology, 42,* 377–399.

Gerard, H. (1983). School desegregation: The social science role. *American Psychologist, 38,* 869–878.

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity.* New York: Touchstone.

Grier, W. H., & Coobs, P. M. (1968). *Black rage.* New York: Basic Books.

Hare, B. R., & Costenell, L. A. (1985). No place to run, no place to hide: Comparative status and future prospects of Black boys. In M. B. Spencer, G. K. Brookins, & W. Allen (Eds.), *Beginnings: The social and affective development of Black children* (pp. 201–214). Hillsdale, NJ: Erlbaum.

Herrnstein, R. A., & Murray, C. (1994). *The bell curve.* New York: Grove Press.

Hewitt, N. M., & Seymour, E. (1991). *Factors contributing to high attrition rates among science and engineering undergraduate majors.* Unpublished report to the Alfred P. Sloan Foundation.

Howard, J., & Hammond, R. (1985, September 9). Rumors of inferiority. *New Republic, 72,* 18–23.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107,* 139–155.

James, W. (1950). *The principles of psychology* (Vol. 1). New York: Dover. (Original work published 1890)

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1–123.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Johnides, J., von Hippel, W., Lerner, J. S., & Nagda, B. (1992, August). *Evaluation of minority retention programs: The undergraduate research opportunities program at the University of Michigan.* Paper presented at the 100th Annual Convention of the American Psychological Association, Washington, DC.

Jones, E. E., Farina, A., Hastorf, A. H., Markus, H., Miller, O. T., & Scott, R. A. (1984). *Social stigma: The psychology of marked relationships.* New York: Freeman.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.

Jones, J. M. (1997). *Prejudice and racism* (2nd ed.). New York: McGraw-Hill.

Kamin, L. (1974). *The science and politics of I.Q.* Hillsdale, NJ: Erlbaum.

Kardiner, A., & Ovesey, L. (1951). *The mark of oppression: Explorations in the personality of the American Negro.* New York: Norton.

Kleck, R. E., & Strenta, A. (1980). Perceptions of the impact of negatively valued physical characteristics on social interactions. *Journal of Personality and Social Psychology, 39,* 861–873.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J.-L. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 75–104). Hillsdale, NJ: Erlbaum.

Levin, J., & Wyckoff, J. (1988). Effective advising: Identifying students most likely to persist and succeed in engineering. *Engineering Education, 78,* 178–182.

Lewin, K. (1941). *Resolving social conflict.* New York: Harper & Row.

Linn, M.C. (1994). The tyranny of the mean: Gender and expectations. *Notices of the American Mathematical Society, 41,* 766–769.

Lord, C. G., & Saenz, D. S. (1985). Memory deficits and memory surfeits: Differential cognitive consequences of tokenism for tokens and observers. *Journal of Personality and Social Psychology, 49,* 918–926.

Lovely, R. (1987, February). *Selection of undergraduate majors by high ability students: Sex difference and attrition of science majors.* Paper presented at the annual meeting of the Association for the Study of Higher Education, San Diego, CA.

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (in press). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin.*

Mead, G. H. (1934). *Mind, self, and society.* Chicago: University of Chicago Press.

Miller, L. S. (1995). *An American imperative: Accelerating minority educational advancement.* New Haven, CT: Yale University Press.

Miller, L. S. (1996, March). *Promoting high academic achievement among non-Asian minorities.* Paper presented at the Princeton University Conference on Higher Education, Princeton, NJ.

National Center for Educational Statistics. (1992). *National Educational Longitudinal Study of 1988: First follow-up. Student component data file user's manual.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Coalition of Advocates for Students Report. (1988, December 12). *The Ann Arbor News,* pp. A1, A4.

Nettles, M. T. (1988). *Toward undergraduate student equality in American higher education.* New York: Greenwood.

Ogbu, J. (1986). The consequences of the American caste system. In U. Neisser (Ed.), *The school achievement of minority children: New perspectives* (pp. 19–56). Hillsdale, NJ: Erlbaum.

Osborne, J. (1994). Academics, self-esteem, and race: A look at the underlying assumption of the disidentification hypothesis. *Personality and Social Psychology Bulletin, 21,* 449–455.

Porter, J. R., & Washington, R. E. (1979). Black identity and self-esteem: A review of the studies of Black self-concept, 1968–1978. *Annual Review of Sociology, 5,* 53–74.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic groups* (College Board Report No. 93-1, ETS No. 94.27). New York: College Entrance Examination Board.

Rosenberg, M. (1979). *Conceiving self.* New York: Basic Books.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 337–384). New York: Academic Press.

Sarason, I. G. (1980). Introduction to the study of test anxiety. In I. G. Sarason (Ed.), *Test anxiety: Theory, research, and applications* (pp. 57–78). Hillsdale, NJ: Erlbaum.

Sartre, J. P. (1965). *Anti-Semite and Jew.* New York: Schocken Books. (Original work published 1946)

Schlenker, B. R., & Weigold, M. F. (1989). Goals and the self-identification process: Constructing desired identities. In L. A. Pervin (Ed.), *Goals concepts in personality and social psychology* (pp. 243–290). Hillsdale, NJ: Erlbaum.

Simmons, R. G., Brown, L., Bush, D. M., & Blyth, D. A. (1978). Self-esteem and achievement of Black and White adolescents. *Social Problems, 26,* 86–96.

Solomon, R. P. (1992). *Forging a separatist culture.* Albany: State University of New York Press.

Spencer, S., Steele, C. M., & Quinn, D. (1997). *Under suspicion of inability: Stereotype threat and women's math performance.* Manuscript submitted for publication.

Steele, C. M. (1975). Name-calling and compliance. *Journal of Personality and Social Psychology, 31,* 361–369.

Steele, C. M. (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). New York: Academic Press.

Steele, C. M. (1992, April). Race and the schooling of Black Americans. *The Atlantic Monthly,* pp. 68–78.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797–811.

Steele, S. (1990). *The content of our character.* New York: St. Martin's Press.

Strenta, A. C., Elliott, R., Adair, R., Scott, J., & Matier, M. (1993). *Choosing and leaving science in highly selective institutions.* Unpublished report to the Alfred P. Sloan Foundation.

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 181–227). New York: Academic Press.

Treisman, U. (1985). *A study of mathematics performance of Black students at the University of California, Berkeley.* Unpublished report.

Vars, F. E., & Bowen, W. G. (1997). *SAT scores, race, and academic performance: New evidence from academically successful colleges.* Unpublished manuscript.

Ware, N. C., & Dill, D. (1986, March). *Persistence in science among mathematically able male and female college students with pre-college plans for a scientific major.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Ware, N. C., Steckler, N. A., & Leserman, J. (1985). Undergraduate women: Who chooses a science major? *Journal of Higher Education, 56,* 73–84.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76,* 92–104.

Wylie, R. (1979). *The self-concept* (Vol. 2). Lincoln: University of Nebraska Press.

Reducing Prejudice: Combating Intergroup Biases

John F. Dovidio[1] and Samuel L. Gaertner

Department of Psychology, Colgate University, Hamilton, New York (J.F.D.), and Department of Psychology, University of Delaware, Newark, Delaware (S.L.G.)

Abstract

Strategies for reducing prejudice may be directed at the traditional, intentional form of prejudice or at more subtle and perhaps less conscious contemporary forms. Whereas the traditional form of prejudice may be reduced by direct educational and attitude-change techniques, contemporary forms may require alternative strategies oriented toward the individual or involving intergroup contact. Individual-oriented techniques can involve leading people who possess contemporary prejudices to discover inconsistencies among their self-images, values, and behaviors; such inconsistencies can arouse negative emotional states (e.g., guilt), which motivate the development of more favorable attitudes. Intergroup strategies can involve structuring intergroup contact to produce more individualized perceptions of the members of the other group, foster personalized interactions between members of the different groups, or redefine group boundaries to create more inclusive, superordinate representations of the groups. Understanding the nature and bases of prejudice can thus guide, theoretically and pragmatically, interventions that can effectively reduce both traditional and contemporary forms of prejudice.

Keywords

attitude change; intergroup contact; prejudice; racism; social categorization

Prejudice is commonly defined as an unfair negative attitude toward a social group or a member of that group. Stereotypes, which are overgeneralizations about a group or its members that are factually incorrect and inordinately rigid, are a set of beliefs that can accompany the negative feelings associated with prejudice. Traditional approaches consider prejudice, like other attitudes, to be acquired through socialization and supported by the beliefs, attitudes, and values of friends and peer groups (see Jones, 1997). We consider the nature of traditional and contemporary forms of prejudice, particularly racial prejudice, and review a range of techniques that have been demonstrated empirically to reduce prejudice and other forms of intergroup bias. Bias can occur in many forms, and thus it has been assessed by a range of measures. These measures include standardized tests of prejudice toward another social group, stereotypes, evaluations of and feelings about specific group members and about the group in general, support for policies and individual actions benefiting the other group, and interaction and friendship patterns.

In part because of changing norms and the Civil Rights Act and other legislative interventions that made discrimination not simply immoral but also illegal, overt expressions of prejudice have declined significantly over the past 35 years. Contemporary forms of prejudice, however, continue to exist and affect the lives of people in subtle but significant ways (Dovidio & Gaertner, 1998; Gaertner & Dovidio, 1986). The negative feelings and beliefs that underlie contemporary forms of prejudice may be rooted in either individual processes (such as cognitive and motivational biases and socialization) or intergroup processes (such as realistic group conflict or biases associated with the mere categorization of people into in-groups and out-groups). These negative biases may occur spontaneously, automatically, and without full awareness.

Many contemporary approaches to prejudice based on race, ethnicity, or sex acknowledge the persistence of overt, intentional forms of prejudice but also consider the role of these automatic or unconscious processes[2] and the consequent indirect expressions of bias. With respect to the racial prejudice of white Americans toward blacks, for example, in contrast to "old-fashioned" racism, which is blatant, aversive racism represents a subtle, often unintentional, form of bias that characterizes many white Americans who possess strong egalitarian values and who believe that they are nonprejudiced. Aversive racists also possess negative racial feelings and beliefs (which develop through normal socialization or reflect social-categorization biases) that they are unaware of or that they try to dissociate from their nonprejudiced self-images. Because aversive racists consciously endorse egalitarian values, they will not discriminate directly and openly in ways that can be attributed to racism; however, because of their negative feelings, they will discriminate, often unintentionally, when their behavior can be justified on the basis of some factor other than race

Copyright © 1999 American Psychological Society

(e.g., questionable qualifications for a position). Thus, aversive racists may regularly engage in discrimination while they maintain self-images of being nonprejudiced. According to symbolic racism theory, a related perspective that has emphasized the role of politically conservative rather than liberal ideology (Sears, 1988), negative feelings toward blacks that whites acquire early in life persist into adulthood but are expressed indirectly and symbolically, in terms of opposition to busing or resistance to preferential treatment, rather than directly or overtly, as in support for segregation.

Contemporary expressions of bias may also reflect a dissociation between cultural stereotypes, which develop through common socialization experiences and because of repeated exposure generally become automatically activated, and individual differences in prejudicial motivations. Although whites both high and low in prejudice may be equally aware of cultural stereotypes and show similar levels of automatic activation, only those low in prejudice make a conscious attempt to prevent those negative stereotypes from influencing their behavior (Devine & Monteith, 1993).

INDIVIDUAL PROCESSES AND PREJUDICE REDUCTION

Attempts to reduce the direct, traditional form of racial prejudice typically involve educational strategies to enhance knowledge and appreciation of other groups (e.g., multicultural education programs), emphasize norms that prejudice is wrong, and involve direct persuasive strategies (e.g., mass media appeals) or indirect attitude-change techniques that make people aware of inconsistencies in their attitudes and behaviors (Stephan & Stephan, 1984). Other techniques are aimed at changing or diluting stereotypes by presenting counterstereotypic or nonstereotypic information about group members. Providing stereotype-disconfirming information is more effective when the information concerns a broad range of group members who are otherwise typical of their group rather than when the information concerns a single person who is not a prototypical representative of the group. In the latter case, people are likely to maintain their overall stereotype of the group while subtyping, with another stereotype, group members who disconfirm the general group stereotype (e.g., black athletes; Hewstone, 1996). The effectiveness of multicultural education programs is supported by the results of controlled intervention programs in the real world; evidence of the effectiveness of attitude- and stereotype-change approaches, and the hypothesized underlying processes, comes largely (but not exclusively) from experimental laboratory research.

Approaches for dealing with the traditional form of prejudice are generally less effective for combating the contemporary forms. With respect to contemporary racism, for example, whites already consciously endorse egalitarian, nonprejudiced views and disavow traditional stereotypes. Instead, indirect strategies that benefit from people's genuine motivation to be nonprejudiced may be more effective for reducing contemporary forms of prejudice. For example, techniques that lead people who possess contemporary prejudices to discover inconsistencies among their self-images, values, and behaviors may arouse feelings of guilt, tension about the inconsistencies, or other negative emotional states that can motivate the development of more favorable racial attitudes and produce more favorable intergroup behaviors (even nonverbal behaviors) several months later. Also, people who consciously endorse nonprejudiced attitudes, but whose behaviors may reflect racial bias, commonly experience feelings of guilt and compunction when they become aware of discrepancies between their potential behavior toward minorities (i.e., what they *would* do) and their personal standards (i.e., what they *should* do) during laboratory interventions. These emotional reactions, in turn, can motivate people to control subsequent spontaneous stereotypical responses and behave more favorably in the future (Devine & Monteith, 1993). People's conscious efforts to suppress stereotypically biased reactions can inhibit even the immediate activation of normally automatic associations, and with sufficient practice, these efforts can eliminate automatic stereotype activation over the long term.

Approaches oriented toward the individual, however, are not the only way to combat contemporary forms of prejudice. Strategies that emphasize intergroup processes, such as intergroup contact and social categorization and identity, are alternative, complementary approaches.

INTERGROUP CONTACT

Real-world interventions, laboratory studies, and survey studies have demonstrated that intergroup contact under specified conditions (including equal status between the groups, cooperative intergroup interactions, opportunities for personal acquaintance, and supportive egalitarian norms) is a powerful technique for reducing intergroup bias and conflict (Pettigrew, 1998). Drawing on these principles, cooperative learning and "jigsaw"

classroom interventions (Aronson & Patnoe, 1997) are designed to increase interdependence between members of different groups working on a designated problem-solving task and to enhance appreciation for the resources they bring to the task. Cooperation is effective for reducing subsequent intergroup bias when the task is completed successfully, group contributions to solving the problem are seen as different or complementary, and the interaction among participants during the task is friendly, personal, and supportive.

Recent research has attempted to elucidate how the different factors of intergroup contact (e.g., cooperation, personal interaction) operate to reduce bias. Engaging in activities to achieve common, superordinate goals, for instance, changes the functional relations between groups from actual or symbolic competition to cooperation. Through psychological processes to restore cognitive balance or reduce inconsistency between actions and attitudes, attitudes toward members of the other group and toward the group as a whole may improve to be consistent with the positive nature of the interaction. Also, the rewarding properties of achieving success may become associated with members of other groups, thereby increasing attraction.

SOCIAL CATEGORIZATION AND IDENTITY

Factors of intergroup contact, such as cooperation, may also reduce bias through reducing the salience of the intergroup boundaries, that is, through *decategorization.* According to this perspective, interaction during intergroup contact can individuate members of the out-group by revealing variability in their opinions (Wilder, 1986) or can produce interactions in which people are seen as unique individuals (personalization), with the exchange of intimate information (Brewer & Miller, 1984). Alternatively, intergroup contact may be structured to maintain but alter the nature of group boundaries, that is, to produce *recategorization.* One recategorization approach involves either creating or increasing the salience of crosscutting group memberships. Making interactants aware that members of another group are also members of one's own group when groups are defined by a different dimension can improve intergroup attitudes (Urban & Miller, 1998). Another recategorization strategy, represented by our own work on the Common In-Group Identity Model, involves interventions to change people's conceptions of groups, so that they think of membership not in terms of several different groups, but in terms of one, more inclusive group (Gaertner, Dovidio, Anastasio, Bachman, & Rust, 1993).

The Common In-Group Identity Model recognizes the central role of social categorization in reducing as well as in creating intergroup bias (Tajfel & Turner, 1979). Specifically, if members of different groups are induced to conceive of themselves more as members of a single, superordinate group rather than as members of two separate groups, attitudes toward former out-group members will become more positive through processes involving pro-in-group bias. Thus, changing the basis of categorization from race to an alternative dimension can alter who is a "we" and who is a "they," undermining a contributing force to contemporary forms of racism, such as aversive racism. The development of a superordinate identity does not always require people to abandon their previous group identities; they may possess dual identities, conceiving of themselves as belonging both to the superordinate group and to one of the original two groups included within the new, larger group. The model also recognizes that decategorization (seeing people as separate individuals) can also reduce bias. In contrast, perceptions of the groups as different entities (we/they) maintains and reinforces bias. The Common In-Group Identity Model is presented schematically in Figure 1.

In experiments in the laboratory and in the field, and in surveys in natural settings (a multi-ethnic high school, banking mergers, and blended families), we have found evidence consistent with the Common In-Group Identity Model and the hypothesis that intergroup contact can reduce prejudice. Specifically, we have found that key aspects of intergroup contact, such as cooperation, decrease intergroup bias in part through changing cognitive representations of the groups. The development of a common in-group identity also facilitates helping behaviors and self-disclosing interactions that can produce reciprocally positive responses and that can further reduce intergroup prejudices through other mechanisms such as personalization.

Moreover, the development of a common in-group identity does not necessarily require groups to forsake their original identities. Threats to important personal identities or the "positive distinctiveness" of one's group can, in fact, exacerbate intergroup prejudices. The development of a dual identity (two subgroups in one group; see Fig. 1), in which original and superordinate group memberships are simultaneously salient, is explicitly considered in the model. Even when racial or ethnic identity is strong, perceptions of a superordinate connection enhance interracial trust and acceptance. Indeed, the development of a dual identity, in terms of a bicultural or multicultural identity, not only is possible but can contribute to the social

Copyright © 1999 American Psychological Society

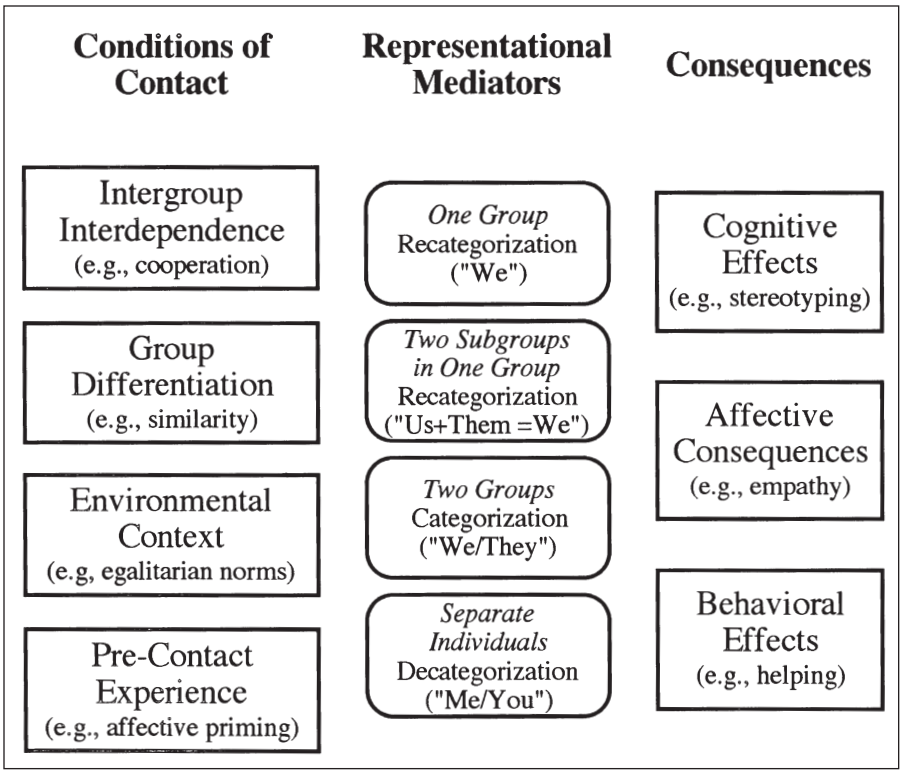


Fig. 1. The Common In-Group Identity Model. In this model, elements of an intergroup contact situation (e.g., intergroup interdependence) influence cognitive representations of the groups as one superordinate group (recategorization), as two subgroups in one group (recategorization involving a dual identity), as two groups (categorization), or as separate individuals (decategorization). Recategorization and decategorization, in turn, can both reduce cognitive, affective, and behavioral biases, but in different ways. Recategorization reduces bias by extending the benefits of in-group favoritism to former out-group members. Attitudes and behavior toward these former out-group members thus become more favorable, approaching attitudes and behaviors toward in-group members. Decategorization, in contrast, reduces favoritism toward original in-group members as they become perceived as separate individuals rather than members of one's own group.

adjustment, psychological adaptation, and overall well-being of minority-group members (LaFromboise, Coleman, & Gerton, 1993). Recognizing both different and common group membership, a more complex form of a common in-group identity, may also increase the generalizability of the benefits of intergroup contact for prejudice reduction. The development of a common in-group identity contributes to more positive attitudes toward members of other groups present in the contact situation, whereas recognition of the separate group memberships provides the associative link by which these more positive attitudes may generalize to other members of the groups not directly involved in the contact situation.

CONCLUSION

Prejudice can occur in its blatant, traditional form, or it may be rooted in unconscious and automatic negative feelings and beliefs that characterize contemporary forms. Whereas the traditional form of prejudice may be combated by using direct techniques involving attitude change and education, addressing contemporary forms requires alternative strategies. Individual-level strategies engage the genuine motivations of people to be nonprejudiced. Intergroup approaches focus on realistic group conflict or the psychological effects of categorizing people into in-groups and out-groups. The benefits of intergroup contact can occur through many routes, such as producing more individuated perceptions of out-group members and more personalized relationships. Intergroup contact can also produce more inclusive, superordinate representations of the groups, which can harness the psychological forces that contribute to intergroup bias and redirect them to improve attitudes toward people who would otherwise be recognized only as out-group members. Understanding the processes involved in the nature and development of prejudice can thus guide, both theoretically and pragmatically, interventions that can effectively reduce both traditional and contemporary forms of prejudice.

Recommended Reading

Brewer, M.B., & Miller, N. (1996). *Intergroup relations.* Pacific Grove, CA: Brooks/Cole.
Brown, R.J. (1995). *Prejudice.* Cambridge, MA: Blackwell.
Hawley, W.D., & Jackson, A.W. (Eds.). (1995). *Toward a common destiny: Improving race and ethnic relations in America.* San Francisco: Jossey-Bass.
Landis, D., & Bhagat, R.S. (Eds.). (1996). *Handbook of intercultural training.* Thousand Oaks, CA: Sage.
Stephan, W.G., & Stephan, C.W. (1996). *Intergroup relations.* Boulder, CO: Westview Press.

Acknowledgments—Preparation of this article was facilitated by National Institute of Mental Health Grant MH 48721.

Notes

1. Address correspondence to John F. Dovidio, Department of Psychology, Colgate University, Hamilton, NY 13346; e-mail: jdovidio@mail.colgate.edu.

2. For further information and a demonstration in which you can test the automaticity of your own racial attitudes using the Implicit Association Test, see Anthony Greenwald's World Wide Web site: http://weber.u.washington.edu/~agg/ (e-mail: agg@u.washington.edu).

References

Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom*. New York: Longman.

Brewer, M.B., & Miller, N. (1984). Beyond the contact hypothesis: Theoretical perspectives on desegregation. In N. Miller & M.B. Brewer (Eds.), *Groups in contact: The psychology of desegregation* (pp. 281–302). Orlando, FL: Academic Press.

Devine, P.G., & Monteith, M.J. (1993). The role of discrepancy-associated affect in prejudice reduction. In D.M. Mackie & D.L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in intergroup perception* (pp. 317–344). Orlando, FL: Academic Press.

Dovidio, J.F., & Gaertner, S.L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. Eberhardt & S.T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 3–32). Newbury Park, CA: Sage.

Gaertner, S.L., & Dovidio, J.F. (1986). The aversive form of racism. In J.F. Dovidio & S.L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Orlando, FL: Academic Press.

Gaertner, S.L., Dovidio, J.F., Anastasio, P.A., Bachman, B.A., & Rust, M.C. (1993). The Common Ingroup Identity Model: Recategorization and the reduction of intergroup bias. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 4, pp. 1–26). London: Wiley.

Hewstone, M. (1996). Contact and categorization: Social psychological interventions to change intergroup relations. In N. Macrae, M. Hewstone, & C. Stangor (Eds.), *Foundations of stereotypes and stereotyping* (pp. 323–368). New York: Guilford Press.

Jones, J.M. (1997). *Prejudice and racism* (2nd ed.). New York: McGraw-Hill.

LaFromboise, T., Coleman, H.L.K., & Gerton, J. (1993). Psychological impact of biculturalism: Evidence and theory. *Psychological Bulletin, 114,* 395–412.

Pettigrew, T.F. (1998). Intergroup Contact Theory. *Annual Review of Psychology, 49,* 65–85.

Sears, D.O. (1988). Symbolic racism. In P.A. Katz & D.A. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (pp. 53–84). New York: Plenum Press.

Stephan, W.G., & Stephan, C.W. (1984). The role of ignorance in intergroup relations. In N. Miller & M.B. Brewer (Eds.), *Groups in contact: The psychology of desegregation* (pp. 229–257). Orlando, FL: Academic Press.

Tajfel, H., & Turner, J.C. (1979). An integrative theory of intergroup conflict. In W.G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Monterey, CA: Brooks/Cole.

Urban, L.M., & Miller, N. (1998). A theoretical analysis of crossed categorization effects: A meta-analysis. *Journal of Personality and Social Psychology, 74,* 894–908.

Wilder, D.A. (1986). Social categorization: Implications for creation and reduction of intergroup bias. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 291–355). Orlando, FL: Academic Press.

Copyright © 1999 American Psychological Society

ADVANCE

EXCELLENCE | EQUITY | DIVERSITY

AVOIDING BIAS IN READING & WRITING EVALUATIONS

evaluations may differ systematically based on gender or race/ethnicity

***Descriptive words* may be used differently in evaluating members of different social groups**:

- **Grindstone adjectives** - (*hardworking, conscientious, dependable, meticulous, thorough, diligent, dedicated, careful*) – are sometimes used more for women, implying that women succeed more through effort than ability.
- **Ability traits** – *talented, smart, able, capable, brilliant*
- **Standout adjectives** – *best, superior, excellent*

***Descriptive phrases* can unintentionally influence a reader.**

1. **Using first names for women** or minority faculty and titles for men (*Joan* was an asset to our department." –vs.- "*Dr. Smith* was an asset to our department.")

2. **Gendered adjectives** ("Dr. Sarah Gray is a *caring, compassionate* physician" –vs. – Dr. Joel Gray has been very *successful with his patients*")

3. **Doubt raisers** or negative language ("although her publications are not numerous" or "while not the best student I have had, s/he")

4. **Potentially negative language** ("S/he requires only minimal supervision" or "S/he is totally intolerant of shoddy research")

5. **Faint praise** ("S/he worked hard on projects that s/he was assigned" or "S/he has never had temper tantrums")

6. **Hedges** ("S/he responds well to feedback")

7. **Irrelevancy** ("S/he is an avid skier and stamp collector")

8. **Unnecessarily invoking a stereotype** ("She is not overly emotional"; "He is very confident yet not arrogant"; or "S/he is extremely productive, especially as someone who attended inner city schools and a large state university"

Based on: Trix, F. & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. Discourse & Society 14: 191-220; Wysocki, V. & Schmader, T. (2005). Recommendation letters of faculty candidates for positions in Chemistry and Biochemistry at UA; and Women in Science & Engineering Leadership Initiative (WISELI), University of Wisconsin, Madison, Handout on Reviewing applicants: Research on Bias and Assumptions.

THE DUAL-CAREER-COUPLE PROBLEM

A type of exclusion principle sometimes operates for married physicists, allowing only one spouse to occupy a job in a given location. Couples are trying to find ways to stay in their chosen fields—and in their marriages.

Laurie McNeil and Marc Sher

Physicists are increasingly faced with a difficult two-body problem—the challenge of finding two professional jobs, possibly two physics jobs, in the same geographic area. More women than ever are entering the physics profession, and they are preferentially marrying scientists. Naturally, they are seeking employment in the same region as their spouses. Yet, in most locales, it's difficult enough to find one physics position, much less a pair of them. If one of the two partners fails to find suitable employment, he or she may be forced to forgo a career in physics.

The dual-career problem affects institutions as well as job seekers. To attract a strongly recruited candidate, employers are under pressure to find a position for the partner. Even if a suitable job is available, the institution may prefer to conduct a full search to fill it and not spend it on the partner. But if the employer does not find a reasonable accommodation for the spouse, the candidate may turn down the proffered position or leave after only a few years.

Despite such difficulties, the two-body problem is not totally intractable. Satisfactory and creative solutions can and have been found. To uncover such solutions, in 1998 we conducted a web-based survey of physicists whose spouses or partners are scientists. We summarize here the outcome of that survey.[1]

The impact on women

In the three decades since 1966, the percentage of women earning PhD's in physics each year has grown[2] from around 2% to roughly 13%. Currently, women constitute only 6% of US physicists overall, but they make up about 13% of all physicists under the age of 31.[3] About half of all women physicists are married, compared to 74% of all male physicists. The figure on the opposite page shows that the female physicists are far more likely to be married to physicists and other scientists than are male physicists.[4]

LAURIE MCNEIL *is a professor at the University of North Carolina, Chapel Hill, both in the department of physics and astronomy and in the curriculum in applied and materials sciences.* MARC SHER *is a professor of physics at the College of William and Mary.*

Thus the dual-career-couple problem has a disproportionate effect on women. Although statistics are difficult to obtain, anecdotal evidence suggests that dual-career employment problems may cause some women to leave physics altogether.

Our survey gathered data on the extent of the problem and possible solutions to it. We queried physicists whose partners are scientists about their experiences in finding employment and about the kinds of positions they had accepted. Our primary goal was to obtain information about possible approaches to the problem; we did not attempt to use rigorous statistical sampling techniques or sophisticated quantitative analysis of the responses.

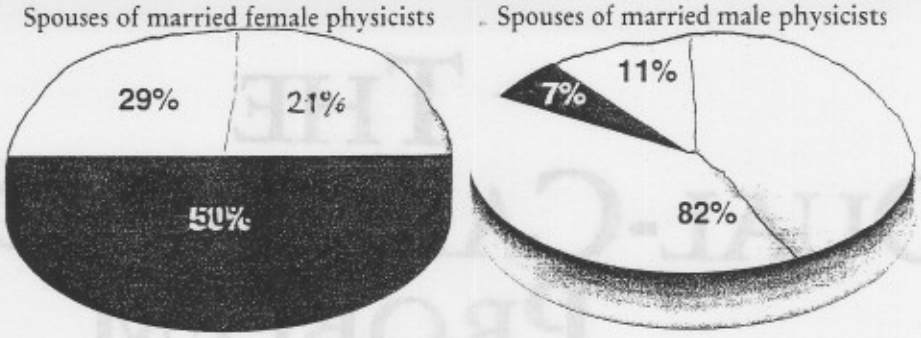
We recognize that the dual-career-couple situation is a problem throughout society and is neither restricted to academia nor to the scientific community. We focused on physics, however, as being more severely affected because of the large gender disparity there, as well as the low density of physics positions in most geographic areas.

Our survey contained 26 questions, 8 of which required narrative answers.[1] In addition to obtaining the demographics of our respondents and their partners—such as gender, age, field of training and type of position held—we asked open-ended questions about the problems that they (and others) had faced, institutional responses, and various employment arrangements.

We arranged through the American Physical Society (APS) to send e-mail messages in January 1998 to the mailing lists of several relevant APS subunits, including two with a high proportion of female subscribers. With funding from the APS Committee on the Status of Women in Physics, we sent flyers to every college and university physics department in the country, and we posted notices in PHYSICS TODAY and *APS News*.

We received 632 replies, 57% from women (not all of them physicists) and 43% from men. We worded the survey in such a way that only one member of a given couple would respond, to avoid double counting. Of the respondents, 89% had partners who were scientists; 45% were physicists married to other physicists. Thus, we tapped, at the very least, 34% of an estimated 830 female physicists who are married to scientists. We take the strong response to indicate the high level of concern about the dual-career problem.

© 1999 American Institute of Physics, S-0031-9228-9907-030-1



SPOUSES OF MARRIED PHYSICISTS. a About half of all female physicists are married, and half of those are married to fellow physicists (red area). Another 29% are wed to scientists other than physicists (yellow), and 21% to nonscientists (blue). b Among the 74% of male physicists who are married, 7% of their spouses are physicists, 11% are scientists other than physicists, and 82% are nonscientists. (Data from ref. 4.)

The mean age of the female respondents was 37.2 years; that of the males was 40.1. About 32% of our respondents had found faculty jobs, 13% had accepted positions in industry, 14% were at national laboratories, 13% had landed postdocs and the rest were in other positions. Just over 50% of the couples surveyed reported that children had affected their job choices, 34% said they had not, and 15% told us they had no children and did not plan to have any. Of much greater importance to us than the statistics, however, were the detailed narrative responses given by a large number of our respondents (see the boxes on pages 34 and 35).

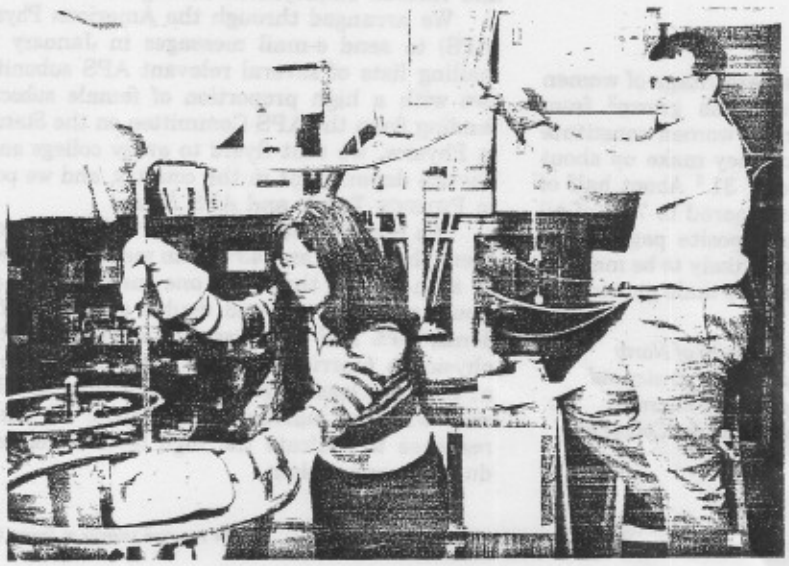
Settling for less

In the ideal job search, of course, both partners find jobs well suited to their qualifications in the same location. But the real outcome is often far different. Our survey found that 60% of all respondents or their partners "had to take a lower-level science or non-scientific job (or no job) in their most recent job search," in the words of the survey.

Some couples choose to have one spouse play the role of "leading partner" and take the best job available. The "trailing partner" then tries to find a suitable job in the same location. The choice of who will take on which role is determined or influenced by several factors, such as professional seniority, research specialty, preference in employment type, personal dynamics, or income opportunity. Relative age can also be a factor, inasmuch as the older person is often further along the career path. Using the survey data, we calculated the average age difference between the respondents and their spouses. The spread was 2.1 years, with the male being the older partner 85% of the time. If age is a major factor in determining the leading partner in the job search, the woman will more often be the trailing spouse. Regardless of which member leads or why, the trailing partner is often hard-pressed to find suitable employment; he or she may end up underemployed or unemployed, and may leave physics altogether.

Possible solutions

Although, to a degree, the dual-career problem is a personal one, which couples must resolve in their own ways, the solutions are inevitably influenced by the institutions where one or both partners are seeking employment. The response of potential employers can either ease or exacerbate the dual-career problem. In the responses to our survey, we have collected examples of both. On the negative side (see the box on page 34), our respondents reported that some institutions actually give reduced consideration to dual-career couples, claiming that it is easier to hire and retain people who are unencumbered, and that some others simply ignore the problem. Some institutions take advantage of the spouse of a staff member they have hired; such "captive spouses" are often geographically trapped and can thus be offered lower-level positions. Other institutions simply refuse to consider the spouse of an existing employee, often citing antinepotism rules (such rules are of questionable



SHARING THE WORK. Natalie Adolphi and Andrew McDowell need both sets of hands on the transfer tube to refill the superconducting magnet in their lab with liquid helium. As tenure-track assistant professors at Knox College, the two share a full-time equivalent position as well as research grants and the lab in which they are pictured. They use solid-state nuclear magnetic resonance to study quasicrystalline metals that absorb hydrogen. At home, they share the responsibility of raising their five-year-old daughter.

Dual-Career Job Hunting: The Down Side

A number of respondents to our web-based survey shared with us some of the negative experiences they had had during their job searches.

▷"The department chair called me at home and asked me several questions about my marital status. He said that he knew these were illegal questions but that he was going to ask them anyway and I could decline to answer them if I wanted. When he found out I was married to a physicist, he said there would be no opportunities for him to be employed in the area. He also said they now screen all candidates because they have offered jobs many times only to be turned down in the end because a spouse could not find a job. A week later I called and found out I was totally off the list. I reported this to the dean and the search was canceled."

▷"In several places I was asked, even before the actual interview, if I would consider accepting the position even if my husband did not get an offer. My husband, on the other hand, never got this kind of question."

▷"I was told that they had already decided not to pursue my application because they 'knew' that I wouldn't be interested in moving since my husband wasn't moving to a position in the area."

▷"My manager had felt that it wouldn't be so difficult for my husband to find a job there; he commented that his wife (a stenographer) had been able to find jobs anywhere."

▷"I remember in particular one senior male faculty member telling me how hard it is to get new professors, because so many of them had spouses who were scientists. This faculty member said he was not about to 'burn' a tenure slot just for somebody's spouse."

▷"One cited antinepotism rules as making it impossible to consider both of us (the rules hadn't existed for years, but apparently the department chair was unaware of this fact.)"

▷"When my wife received her first job offer, their physics department had an opening for which I was very well qualified. They acknowledged my qualifications, but refused to interview me because they said they were not interested in my research areas. However within the next two years they hired faculty in closely related fields. I am quite convinced that their perception of me as a 'trailing spouse' affected the decision."

▷"They suggested that I might consider giving up my career."

▷"[I know of] two extremely talented scientists. The husband, a little ahead chronologically in his career, has tenure at a large university. The wife is teaching and doing research at the same university on soft money. Despite her glowing teaching and publication record, she has been constantly passed over on recent job searches. Documents, secretly released to her, seem to indicate the search committee hopes she will just stay, on her soft money 'after all, her husband has tenure. Why waste a real job on her?'"

▷"It is a very bad idea to raise this issue *before* an offer is made, since all negotiating leverage for salary and benefits would be lost. At [three prominent universities], jobs offered to us both as a 'package deal' had miserable salaries as a result of their knowing we wanted to stay together."

▷"One professor suggested to my husband at his interview that one way to solve the two-body problem was to divorce me—not a very sensitive suggestion."

▷"I was told that I should be able to find a lab to work in, as long as I was willing to change fields and didn't expect to be paid; if I 'needed to be paid,' I might be able to teach introductory calculus."

▷"One [interviewer] suggested that I should be available to do 'volunteer' scientific work, because it was my partner's role to support the family."

PARALLEL TRACKS. Janet Tate and David McIntyre sit in front of Memorial Union on the campus of Oregon State University, whose college of science adopted a dual-career couples policy in 1987. Tate and McIntyre—both full-time associate professors of physics—are among 12 couples in which both partners have faculty appointments in science. A further five couples have one partner in science.

legality[1]). On the positive side (see the box on page 35), some hiring institutions informed the trailing spouse of opportunities in the area, helped the person to find full or part-time employment within the same institution, or created various arrangements for the husband and wife to share a position. Some of the kinds of institutional responses we learned about through our survey involve formal responses to the problem, while others are more *ad hoc*. No single strategy works for all dual-career couples, but combinations of strategies can prove useful in particular situations.

Split positions. An increasingly common solution to the two-body problem involves arrangements that are variously called split, shared, or joint positions. They began at colleges and universities in isolated areas, where spouses had few employment opportunities outside the university, but the popularity of such arrangements has been growing rapidly in recent years, and they are now being adopted at some major research universities.

Technically, the terms "split" and "shared" denote different arrangements. In a shared position, two individuals fill a single faculty position, usually on a 50–50 basis. The two occupants of the one position are jointly considered for tenure and promotion (they sink or swim together). The salary and responsibilities are divided evenly. In a split position, a single job is divided into two separate, independent, half-time positions. Each half-position has a separate contract dictating research, teaching, and service duties. Each individual in a half position is separately eligible for tenure, promotion, raises, and benefits. The most serious concern about the shared position, some college administrators told us, is the "all-or-nothing" aspect of the tenure decision: If only one member of the shared position qualifies for tenure, it is awkward to turn down the other. We found that different institutions tend to devise their own combinations of split and shared

THE LONG COMMUTE. Bob and Ruth Howes enjoy a welcome moment together. They have lived in separate states for most of their 33-year marriage. Ruth is the George and Frances Ball Distinguished Professor of Physics and Astronomy at Ball State University, in Muncie, Indiana, while her husband was a professor at the Health Sciences Center of the University of Oklahoma until he recently retired. Ruth quips that moving the children back and forth wasn't bad, but combining families of cats in the summer was really rough.

positions (some are detailed in reference 1), so that here we generically refer to such arrangements as split positions.

Split positions are best for two people in the same field, at roughly the same professional level. In many cases, they provide the only mechanism for both partners to stay active in science, in mainstream positions, and still live together. They have the additional appeal of allowing a couple to balance the demands of both career and family, particularly before they earn tenure or before their children are in school.[5] The extra time can also be used to pursue other research interests. And there is always the hope that the half-time positions may evolve into full-time positions.

An obvious disadvantage of a split position is financial—the total shared salary is no higher than that of a single spouse working full time. Yet the total workload may be greater because each professional is likely to spend more than half time at his or her job. However, at least at colleges and universities, if both partners have research grants with summer support, the total annual salary can reach 1.25 person-years. By picking up extra classes when colleagues go on leave, or by seeking consulting and outside employment, the pair can further increase their annual income.

From the viewpoint of the hiring institution, split positions may require an increased administrative burden, but they often represent a "two for the price of one" package. An academic department or industrial research group can bring in two people, with different approaches, viewpoints, and skills, for little more than a single full-time salary. Even though the partners technically each work only half-time, they are likely to give more than half-time thought and commitment to their jobs. Split positions also provide students with alternative role models. One department chair told us that offering shared positions adds stability to the faculty and reduces departmental difficulties due to attrition.

Dual-Career Job Hunting: The Up Side

Other survey respondents had more favorable experiences.

▷"Since I was offered a postdoc fellowship to work with a particular person, I contacted my potential boss. He, as an individual, responded by giving us names of friends and colleagues to contact about possible jobs. He also contacted some of those people for us to set up exploratory appointments during one of our visits to the university. The director of the department contacted several other department heads on campus to see about appropriate openings. Each time I asked for help, a bit more effort was put into finding my partner an acceptable job. In the end, this paid off, since a position was created for him by the university."

▷"For the first five years I had a part-time job at the university and therefore our department knew about my teaching and research abilities. Thus, when they had an opening for a tenure track position, they hired me."

▷"The institution was very helpful. They spoke with my partner's department to locate a part-time position. This turned into a full-time tenure-track [position] later on."

▷"The university where my husband is a faculty member is on the opposite coast to mine. My employer has provided an ample travel budget in my startup package. This allows me to travel to the national lab where both my husband and I do research. Therefore, I see my husband several times per month at the lab."

▷"They found a temporary salary for him through the University Diversity Funds."

▷"I really liked the fact that all four potential employers treated me as a physicist in my own right and didn't discuss my family situation where it wasn't appropriate."

▷"I felt my present institution was making a good-faith effort to accommodate the situation, and treated my wife like a respected colleague, not a charity case."

▷"We were pleased to find that our applications (for positions at the same institution) were taken seriously and handled equivalently."

▷"Received some assistance from individuals at the facility to which I was applying, in identifying specific job opportunities."

▷"It was incredibly positive to have them work with us from day one to find a position for my spouse. While the first year was only a part-time instructor position, the position has evolved every year and is now a tenure-track chemistry position. The dean was incredibly forthcoming and supportive throughout the process."

▷"It was tremendously helpful to have the admin people set up the interviews. We had already compiled a list of places for my partner to apply, but the position my partner eventually accepted was one that the admin people found on their own."

▷"I was very impressed with the department which ultimately hired us. They already had the details of dual positions worked out, so they could discuss the possibilities directly."

Spousal hiring programs. Split and shared positions are obviously not appropriate when spouses are in different fields, or when they seek different types of employment, such as academic versus industrial. In response to the dual-career-couple problem, a number of universities have established formal spousal hiring programs. These programs help the partners of potential hires to explore the employment possibilities within the univer-

sity, as well as with local companies or institutions. Our survey respondents specifically mentioned such programs at the University of Wisconsin; the University of California, Davis; Purdue University; and the University of Illinois at Urbana-Champaign.

The Spousal Relocation Assistance Program at Purdue, for example, has a half-time relocation specialist who has a comprehensive knowledge of local companies, industries, and organizations, and who will help by providing information to the partner about the area's employment resources and networking possibilities and by alerting appropriate companies and organizations of the talents of the spouse. In addition, the specialist serves as a resource for Purdue deans, vice presidents, directors, and department chairs in their recruitment efforts.

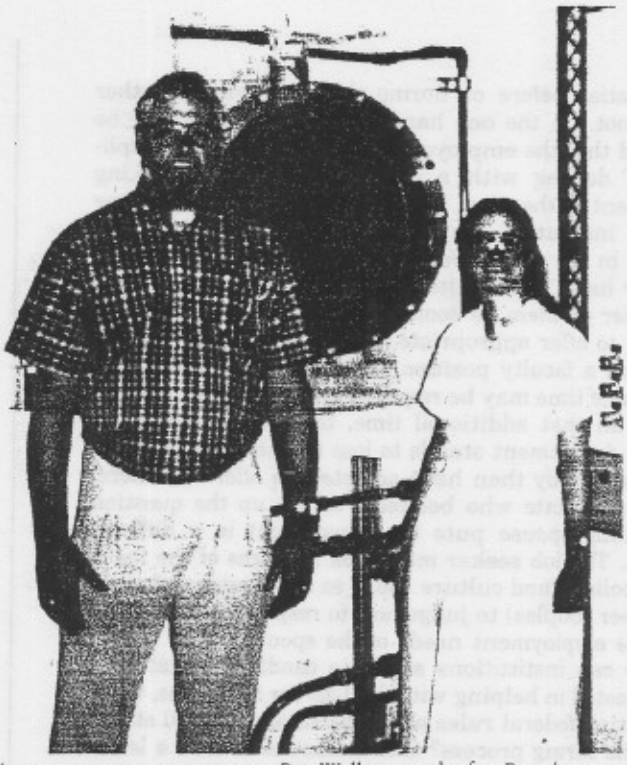
Although department chairs don't have the time and resources to perform the same job as a half-time relocation specialist, they may make useful contacts in the normal course of their job. For example, it is a good idea in any case for physics departments to cultivate close relations with local industries—to learn of job opportunities for the department's own graduates and possibilities for internships, funding, or even collaborative research. These contacts can then be tapped to explore possible openings for the spouses of job candidates.

Alternative academic positions. If no faculty positions are available for a trailing spouse, other possibilities include short-term (2–3 year) postdoctoral research appointments, soft-money research positions, and adjunct or part-time teaching. Our survey found that 13% of the respondents were able to get short-term postdocs. Once such a position ended, the department was sometimes able to find a tenure-track position to avoid losing both partners.

Before embarking on the short-term route, the trailing spouse should inquire about the department's future hiring plans. If the department does not expect to bring in a new person in the trailing spouse's subfield for many years, chances may be dim for the short-term position to turn into a permanent situation. Of course, even if the department *does* expect to hire a person in the given subfield in the next few years, no institution would (or should) be expected to make any promises about that position.

If a tenure-track position does not appear likely, another possibility is a long-term soft-money position, which is largely funded from a grant or contract. Many research groups (especially in high-energy physics) do have long-term positions, which can last for decades. There are also positions involving systems management—one survey respondent reported having gotten a permanent position, doing research half-time and managing the department's server half-time. A 1993 study of soft-money research positions[6] found that 40% of the occupants of such appointments are spouses of tenure-track faculty, and that three quarters of those were women. The occupants of such positions, however, face many disadvantages, including a form of "second-class citizenship."[6]

Adjunct or part-time teaching posts can also be long-term arrangements. They offer advantages for the family with young children. However, the pay is often very low, the adjunct is not considered a full member of the department, and the positions are extremely unstable, requiring the spouse to beg for courses semester by semester. Institutions typically offer little or no support for adjuncts to do research. Anecdotal evidence indicates that adjunct or part-time teaching is the first step on the road out of science for many female scientists. Universities could help improve the morale of valued adjunct professors by giving them contracts, if only for a few years at a time; by possibly allowing nontenure-track instructors to be eligible for teaching awards; by giving adjunct faculty access to institutional resources, such as career counseling; and by finding some departmental funding for travel to conferences.



ACADEMIA AND INDUSTRY. Pat Wallace works for Raychem Corp in Fuquay-Varina, North Carolina, not far from the University of North Carolina, Chapel Hill, where Laurie McNeil is a professor of physics and astronomy and of applied and materials sciences. This couple focused their job search on sites with a university and several industries that employ physicists. They are shown visiting Raychem's main facility in Palo Alto, California. Behind them is the viewing port of an electron accelerator used to process specialty polymer products.

Long-distance commuting. Sometimes, the only way the couple can both continue their careers is to take jobs at locations that are too far apart for a daily commute. We did not specifically ask survey respondents about such long-distance relationships, but we learned from our responses that a number of dual-career couples have spent at least some time living separately. The disadvantages of such an arrangement are obvious, particularly for couples who have, or wish to have, children. Some institutions have softened such hardships by allowing a faculty member to be on leave for one term each year. Many institutions allow sabbaticals every seven years; a number have support for visiting scientists and/or don't object to the occasional year of unpaid leave. With a creative staggering of sabbaticals and leaves, couples can end up spending a rather high percentage of their time together. The increasing use of telecommuting has the potential to alleviate many of the difficulties of dual-career couples, particularly in industry.

A delicate dance

One of the most delicate issues surrounding the two-body problem concerns the appropriate time to discuss the situation with the prospective employer. Half of our survey respondents told a potential employer about the dual-ca-

reer situation before or during the interview; the other half did not. On the one hand, the job candidate may be concerned that the employer will not welcome the complication of dealing with a partner who is also seeking employment in the area. The dual-career couple may fear that the institution may consider their desire to find positions in the same area as a strike against them. On the other hand, the institution may want to know of the dual-career problem as soon as possible so that it can be prepared to offer appropriate, timely assistance. Particularly when a faculty position for the partner is involved, considerable time may be required to put together an offer. If, after all that additional time, the couple rejects the offer, the department stands to lose the next candidate in line, who may by then have accepted an offer elsewhere. Thus, a candidate who belatedly brings up the question of a trailing spouse puts the department in a difficult situation. The job seeker must look for signs of the institution's policies and culture (such as the presence of other dual-career couples) to judge how to respond to a question about the employment needs of the spouse.

How can institutions signal to candidates that they are interested in helping with dual-career situations, without violating federal rules about discussing marital status during the hiring process? One method is to send a letter to all finalists for a position, before the actual interview, informing them of any special dual-career policies that the institution may have, and offering to assist with dual-career issues. The University of Wisconsin—Madison sends such information to tell candidates about its spousal hiring program.

Our hope

As women represent a growing fraction of younger physicists, the number of new hires facing the two-body problem can be expected to increase. It is in the interests of institutions to take an active role in addressing these dual-career situations. Doing so can help them hire and retain the candidates they want and can also reduce the significant barriers experienced by talented women entering the profession. Some of the actions we have described—creative job sharing, spousal hiring programs, part-time appointments—benefit not only the job candidate and the institution, but also the physics profession as a whole. As physicists who have each experienced the dual-career situation directly, we hope that an increasing number of institutions will adopt these measures.

References

1. A detailed report on the survey, sample split-position contracts, spousal hiring programs and links to various resources can be found at http://www.physics.wm.edu/dualcareer.html. The survey report is also on http://www.physics.wm.edu/survey.html.
2. "Career Characteristics of Doctorate Recipients," Committee on Women in Science and Engineering, National Research Council (1999).
3. R. Y. Chu, J. M. Curtin, R. Czujko, *American Physical Society Membership Survey*, American Institute of Physics, New York, New York (1996).
4. P. H. Blondin, A. Benedict, R. Y. Chu, *American Physical Society Membership Survey*, American Institute of Physics, New York, New York (1990).
5. C. B. de Wet, A. P. de Wet, Journal of Women and Minorities in Science and Engineering **3**, 203 (1997); Geotimes **4**, 17 (1995). N. Adolphi, *APS News* December, 1995, p. 5. J. Lubchenco, B. Menge, Biosciences **43**, 243 (1993). L. Katterman, The Scientist, October, 1995, p.16. C. Shrady, Geosciences Canada **21**, 134 (1995).
6. D. Goldberg, A. K. Sakai, Ecological Society of America Bulletin, **74** (2), (1993). ■

HAVE YOU HEARD ABOUT THE NEW PROCESS RGA WITHOUT PUMPS

With Spectra's revolutionary HPQ2 range of high pressure process RGA's monitoring your process tool, you'll be able to take control of your yields.

The ultra compact HPQ2 is used in degas, preclean and process chambers and mounted directly, without the need of valves and pumps, making it much less expensive to install and trouble free in operation.

The HPQ2 runs Spectra's advanced Process Eye software with optional SECS II interface, which integrates with the process tool. And with HPQ2's *Sensor*Log database facility you can access trends and total condition over extended periods.

Take control of your yields. Contact Spectra today, we have the solution to your problem.

SPECTRA
INTERNATIONAL
SENSORS + SUPPORT = SOLUTIONS
www.spectra-sensors.com

USA Tel: +1 408 778 6060 | UK & Europe Tel: +44 (0)1270 250150 | Japan & Asia Tel: +81 (0)7914-6-8061

Circle number 17 on Reader Service Card

Handbook for Faculty Searches and Hiring

Academic Year 2009–10

Please see the following URL for a continually updated version of this document:
http://www.umich.edu/~advproj/handbook.pdf

ADVANCE Handbook for Faculty Searches and Hiring

Table of Contents

I. Introduction

II. Initiating the Search Process

- Composition of the Committee
- Search Committee's Charge
- How to Avoid Having Active Recruitment Efforts Backfire
- Defining the Position
- Language for Announcing Positions
- The Importance of Dual Career Considerations

III. Committee Activity before the Search Begins

- Reviewing the National Pool
- Reviewing Past Departmental Searches

IV. Recruiting Activities during the Search

- Broadening the Pool
- Using Active Recruiting Practices
- Using Active Recruiting Resources
- Creating the Short List

V. Handling Campus Visits

- Candidate Evaluation Sheet

VI. Negotiating the Offer

VII. Getting Off to a Good Start

VIII. Evaluating the Search

Appendix 1: Reading Lists

Appendix 2: Active Recruiting Resources

For more information or additional copies of this resource, please contact the ADVANCE Program at (734) 647-9359 or advanceprogram@umich.edu, or visit the ADVANCE Program's Web site at http://sitemaker.umich.edu/advance.

I. Introduction

Efforts to recruit, retain, and promote diverse faculty in science and engineering have produced slow and uneven results. This has been the case both nationally and at the University of Michigan. Since the summer of 2002, under the auspices of the UM NSF ADVANCE grant, the Strategies and Tactics for Recruiting to Improve Diversity and Excellence (STRIDE) Committee has given presentations to search committees and other interested faculty and administrators aimed at helping with the recruiting and retention of women and other minorities under-represented among the faculty (e.g., racial and ethnic minorities, sexual minorities, people with disabilities). This handbook is designed to integrate and summarize the recruitment and hiring practices that have been identified nationally and by the STRIDE committee as effective, practical, and fair.

The STRIDE committee is composed of a diverse group of senior faculty who are able to advise individuals and departments on hiring practices aimed at increasing both the diversity and excellence of the faculty through presentations, detailed and targeted advice, or focused discussions as needed. Several times a year STRIDE offers a workshop for search committee members and other faculty entitled “Workshop on Faculty Recruitment for Diversity and Excellence.” The PPT of the presentation is accessible at the following URL: http://sitemaker.umich.edu/advance/stride.

After several years of experience with the STRIDE committee’s activities, ADVANCE is able to report real progress in the recruitment of women in each of the three colleges that employ the largest number of scientists and engineers at the University (College of Engineering, LSA Natural Sciences, and Medical School Basic Sciences). As a proportion of science and engineering tenure-track hires, 13% (N=9) of all new hires were women in AY2001 and AY2002 (the “pre-ADVANCE” years), as compared with 31% (N=71) in AY2003–AY2008 (a statistically significant increase).

While many factors no doubt contributed to departments’ willingness and ability to hire more women, STRIDE is the intervention that most directly provided tools and ideas to aid in recruitment.

Moreover, some particular departments have reported especially rapid progress. For example, before the ADVANCE Program, the UM Chemistry Department’s average representation of women in their applicant pool (1998-99 to 2002–03) was 10%. After the ADVANCE Program and the Department’s adoption of “open searches,” the average representation of women in the applicant pool rose to 18%. In the Department of Astronomy, the number of women on the tenure track increased from 0 in AY2001 to 5—or 33%—in AY2006. Both departments—which participated actively in ADVANCE programs and employed recommended hiring practices—have become nationally recognized for the outstanding quality and diversity of their faculty hiring during this period.

The larger context for faculty hiring activities includes both national and federal mandates, state legal constraints, and university commitments. As President Coleman stated in her remarks to the community after the 2006 passage of Proposal 2, "The University of Michigan embraces, promotes, wants, and believes in diversity." As was stated by Laurita Thomas, Associate Vice President for Human Resources, in a letter to the UM community:

The passage of Proposal 2 does not change our commitment, nor does it alter our employment practices or the protections and requirements of various federal and state laws including the Civil Rights Act of 1964, Title IX of the Education Amendments, the Americans

with Disabilities Act, and Michigan's Elliott-Larsen Civil Rights Act, which prohibit a wide array of discrimination extending far beyond issues of race and gender.

We are encouraged to continue to work diligently to recruit and retain the best faculty and staff by creating a community that seeks, welcomes and defends diversity. We will do so in compliance with state and federal laws, and federal law requires that we continue to take affirmative steps (known as affirmative action) in our employment process in order to adhere to the equal employment opportunity and affirmative action provisions of Executive Order 11246 regarding race, gender, color, religion and national origin required of all federal contractors. Proposal 2 specifically states that it does not prohibit actions that are required to establish or maintain eligibility for any federal program, if ineligibility would result in a loss of federal funds to the state. Specifically, this means that:

- The University's nondiscrimination policy remains in full force and effect (see SPG 201.35 http://spg.umich.edu/pdf/201.35.pdf).
- A host of federal and state civil rights laws, including those discussed above, continue to be in effect and applicable to the University.
- The University must continue to adhere to all the requirements of Executive Order 11246.
- As it relates to the employment process, Executive Order 11246 requires all federal contractors, such as U-M, to take affirmative steps to ensure its employment process is fair and equitable and offers equal opportunity in hiring and employment. The types of affirmative steps required include a focus on recruiting and outreach, such as casting the widest net possible when conducting an employment search.
- Executive Order 11246 also requires that federal contractors not discriminate against job applicants or employees.
- The University's standard statement in employment ads, "A Non-Discriminatory/Affirmative Action Employer" or similar language such as "Affirmative Action/Equal Opportunity Employer" is required by Executive Order 11246 and must continue to be used.

Further information regarding the University's nondiscrimination statement, diversity, or affirmative action can be obtained from the Office of Institutional Equity. http://www.hr.umich.edu/oie

II. Initiating the Search Process

The composition of the search committee and its charge are factors likely to have consequences for the outcome of the search. It is important that issues of composition and charge be addressed deliberately and early. STRIDE committee members are happy to meet with department chairs or other decision-makers to help think through issues associated with the composition of, and charge to, the search committee.

Composition of the Committee

- Search committees should include members with different perspectives and expertise, and with a demonstrated commitment to diversity.

• Search committees should include women and underrepresented minorities whenever possible.

• It is often helpful to appoint some search committee members from outside the department. Note, however, that women and minorities are often asked to do significantly more service than majority males, so it is important to keep track of their service load, free them from less significant service tasks, and/or compensate them in other ways.

Initial Discussions of the Search Committee's Charge should:

• verify that its charge includes particular focus on equitable search practices, and the goal of identifying outstanding women and underrepresented minority candidates for the position.

• articulate the fact that diversity and excellence are fully compatible goals and can and should be pursued simultaneously.

• identify selection criteria and develop the position description prior to beginning the search.

• establish plans for actively recruiting women and underrepresented minorities prior to beginning the search.

• review practices that will mitigate the kinds of evaluation biases that social science research has identified that result in unfair evaluations for women and minority candidates.

• include discussion of how the plans to represent the school's or department's commitment to and strategies for hiring and advancing diverse faculty are integrated into the strategies. This may be of particular concern for departments that have few or no women or under-represented minority faculty. In these cases, it may be helpful to develop long-term strategies for recruiting diverse faculty. For example, the department might consider inviting women or minority faculty to give talks and then inviting them to apply for positions the following year.

It may be helpful for the committee to view the videotaped lecture by Professor Virginia Valian, of CUNY, summarizing this research, and discuss it as a group. The lecture may be viewed at the following URL: http://mitworld.mit.edu/video/80

Also examine Professor Valian's interactive tutorial, which can be accessed at the following URL: http://www.hunter.cuny.edu/gendertutorial/tutorials.htm

• remind committee members that STRIDE is available to consult as questions arise throughout the search process.

How to Avoid Having Active Recruitment Efforts Backfire

• Women and minority faculty candidates wish to be evaluated for academic positions on the basis of their scholarly credentials. They will not appreciate subtle or overt indications that they are being valued on other characteristics, such as their gender or race. Women candidates and candidates of color already realize that their gender or race may be a factor in your considerations. It is important that contacts with women and minority candidates for faculty positions focus on their scholarship, qualifications, and potential academic role in the department.

Defining the Position

- Define the position in the widest possible terms consistent with the department's needs. Aim for consensus on specific specialties or requirements, while planning to cast the hiring net as broadly as possible. Make sure that the position description does not needlessly limit the pool of applicants. Some position descriptions may exclude female or minority candidates by focusing too narrowly on subfields in which few specialize.

- Consider as important selection criteria for all candidates (regardless of their own demographic characteristics), the ability of the candidate both to add intellectual diversity to the department, and to work successfully with diverse students and colleagues.

- If women or minority candidates are hired in areas that are not at the center of the department's focus and interest, they may be placed in an unfavorable situation. It is important to carefully think about how the department will support not only the individual, but also the development of that person's area within the department. Consider "cluster hiring," which involves hiring more than one faculty member at a time to work in the same specialization.

- Establish selection criteria and procedures for screening, interviewing candidates, and keeping records before advertising the position.

- Make sure that hiring criteria are directly related to the requirements of the position, clearly understood, and accepted by all members of the committee.

- Get committee consensus on the relative importance of different selection criteria. Plan to create multiple short lists based on different key criteria. (See "Creating the Short List," in section IV, below.)

Language for Announcing Positions

- Proactive language can be included in job descriptions to indicate a department's commitment to diversity. This may make the position more attractive to female and minority candidates. Examples include:

    - "The college is especially interested in qualified candidates who can contribute, through their research, teaching, and/or service, to the diversity and excellence of the academic community."

    - "The University is responsive to the needs of dual career couples."

    - "Women, minorities, individuals with disabilities, and veterans are encouraged to apply."

The Importance of Dual Career Considerations

While it is critical that women and minority candidates be treated first and foremost as the scholars they are, it is equally important that search committees and departments understand the importance of dual career considerations in recruiting women and underrepresented minority faculty in science and engineering. If your search committee and department chair are willing to do their best to help place qualified spouses and partners, you might consider including the following statement in the ads for positions: "The University is responsive to the needs of dual career couples."

At the same time, it is critical that all search committees recognize that it is inappropriate and illegal for individuals' marital or family status to affect evaluation of their application. Knowledge—or guesses—about these matters may not play any role in the committee's

deliberation about candidates' qualifications or the construction of the shortlist. All committee members should recognize this and help maintain a proper focus in committee deliberations, but of course the committee chair has a special responsibility to ensure that the discussion excludes any inappropriate considerations.

The UM Human Resources and Affirmative Action Web site includes a chart comparing legal and discriminatory questions about:

- Family status
- Race
- Religion
- Residence
- Sex
- Age
- Arrests or convictions
- Citizenship or nationality
- Disability

Details can be found at the following URL:
http://www.hr.umich.edu/empserv/department/empsel/legalchart.html

Regardless of candidates' personal characteristics (and without knowing anything about an individual's partner or family status), one feature of the University environment that is likely to be important and attractive to *all* candidates is policies that make it a humane work setting. As you provide that information to all candidates, keep a few notions in mind:

- While it is common for academics to be partnered with other academics, academic *women* are more likely to be partnered with other academics than academic men are. This means that disadvantages that affect two-career academic couples have a disproportionate impact on women.

- At the same time, recognize that there is variability among women in their personal and household circumstances. Do not assume one household type (e.g., a husband and children) applies to all women.

- Make sure everyone on the search committee has a good working knowledge of the UM's dual career support programs. Consult the Provost's Office for further information. Information is also available online at www.provost.umich.edu/programs/pfip.html. This site provides online resources for dual career partners seeking employment. In addition, the document, "University of Michigan Dual Career Program: Roles and Responsibilities & Steps in the Process," a resource for University administrators, is available by contacting the Provost's Office. Precise procedures vary in each school and college, so search committee chairs should consult their department chairs about the correct procedures they should follow.

- Provide all candidates with a copy of the flier, "Dual Career Program at the University of Michigan: A Guide for Prospective and New Faculty Members," which is also available online: www.provost.umich.edu/programs/dual_career/DualCareerBrochure9201.pdf

- Address perceptions that Ann Arbor, as a small city, offers limited opportunities for a candidate's spouse or partner. Make sure candidates know about the diverse employment possibilities their partners might find not only at the university, but also throughout Ann Arbor and in the larger Southeast Michigan area. The Dual Career office can provide

helpful information about Ann Arbor and surrounding communities. (See contact information above).

- Identify someone in the department who can offer to have a confidential conversation (one not to be conveyed to anyone else in the department) with candidates about these issues. This person should be well-informed about all programs supporting faculty members' families, and willing to describe or discuss them with candidates, without transmitting information about the candidate's personal circumstances to the department or the rest of the search committee. Another possibility is to have this person come from outside the interviewing department. For example, the College of Engineering has a committee of senior faculty women who volunteer to serve as contacts for women candidates, and the Associate Dean for Academic Affairs (ADAA) requires that each female candidate meet with a member of this committee.

- If a candidate does mention having a spouse or partner who will need placement help, follow the procedures appropriate in your school or college to arrange interviews or other opportunities for the spouse or partner as early in the hiring process as possible. Your department chair is the best source on this, but it is always possible to get information and assistance from the Dual Career Coordinator in the Provost's office.

The ADVANCE Program can be reached by email at: advanceprogram@umich.edu or by web form request at: http://sitemaker.umich.edu/advance/contact

III. Committee Activity before the Search Begins

The search committee, and/or a larger group in the department, should engage in a relatively extended review of the wider disciplinary context, as well as the department's own past history of searching and hiring, before beginning a new search. The department is more likely to be able to achieve a different outcome from past outcomes if it has some understanding of factors that may have played a role in limiting past success in recruiting women and minorities.

Reviewing the National Pool

- Take steps to identify the national "pools" of qualified candidates for the field as a whole and for subfields in which you are considering hiring. Subfield pools are sometimes quite different from overall pools. ADVANCE Program staff are willing and able to assist you in identifying field and subfield pools.

- Identify any institutions or individuals that are especially successful at producing women and/or under-represented minority doctorates and/or postdoctorates in your field or the desired subfield. Recruit actively from those sources.

Reviewing Past Departmental Searches

- Find out how many women and under-represented minorities have applied for past positions in your department, as a percentage of the total applicant pool.

- Find out how many women and under-represented minorities have been brought to campus for interviews in your field in previous searches.

- If women or under-represented minority candidates have been hired in recent searches, ask the search committees, the department chair, and the recently hired faculty themselves how they were successfully recruited.

- If women or under-represented minority candidates have been offered positions but have turned them down, find out why they have turned them down. ADVANCE staff are willing and able to conduct confidential interviews with such candidates, if you think they might be less than candid in talking with colleagues in the same field. Be sure, in any case, to collect multiple accounts; individual stories often differ. Listen for potential insights into departmental practices that might have been a factor in candidates' decisions. Stories that appear to be highly individual at first may reveal patterns when considered in the aggregate.

The Psychology Department at the University of Michigan successfully recruited faculty of color by maintaining a standing committee to develop information about potential candidates, and following up on that information as opportunities arose.

- Find out what has happened to women and under-represented minorities who were not offered positions in previous searches. Where are they now? Does it appear that something interfered with the assessment of their likely success?

- If no women or under-represented minorities have been offered positions in recent searches, consider redefining departmental evaluation systems in ways that might better take strengths of female and under-represented minority candidates into account. Consider whether positions have been defined too narrowly. If candidates have been ranked on a single list, consider using multiple ranking criteria in the future.

IV. Recruiting Activities during the Search

Broadening the Pool

- Be aware that the University of Michigan's Provost's Faculty Initiative Program (PFIP) provides supplemental resources "to help the schools and colleges and other academic units to hire and retain faculty who contribute to the intellectual diversity of the institution, to assist the dual career partners of tenure track and tenured faculty, and to respond to unique opportunities." This program can often help you recruit and retain women and minority faculty. Consult the Provost's Office for further information.

- View your committee's task as including a process of generating a pool rather than merely tapping it. This may be accomplished by having committee members attend presentations at national meetings and develop a more diverse list of potential future candidates based on those meetings. Candidates identified in this way may be in any field, not necessarily the one targeted for a particular search. In fact, the department may consider creating a committee to generate women and/or minority candidates, who can then be considered for targeted recruitment outside of subfield-defined searches. In addition, the committee may consider issuing promising candidates invitations to visit UM informally to present research before those individuals are ready for an active search. Cultivating future candidates is an important activity for the search committee to undertake, and may require that the search have a longer time horizon than is typical.

- If your department is a significant source of qualified applicants nationally, consider setting aside the traditional constraint against "hiring our own." It may be important, if your department or related ones at UM is a significant producer of the pool, to avoid unduly constraining the search to those trained elsewhere.

- Keep in mind that some eminent universities have only recently begun actively to produce women and minorities Ph.Ds. Therefore, consider candidates from a wide range of institutions.

- Consider the possibility that women and under-represented minorities who have excelled at their research in departments less highly ranked than UM's may be under-placed and might thrive in the University of Michigan research environment.

- Beware of systems of evaluation that inadvertently screen out well-qualified applicants from minority-serving institutions.

- Be careful to place a suitable value on non-traditional career paths. Take into account time spent raising children or getting particular kinds of training, unusual undergraduate degrees, and different job experiences. There is considerable evidence that evaluations of men frequently go up when they have such work experience, while evaluations of women with the same kinds of experience go down.

- Keep in mind that when more than one woman and/or minority candidate is brought in for an interview, women or minority candidates are disproportionately more likely to be hired. Research indicates that interviewers evaluate women and underrepresented minorities more fairly when there is more than one woman in the interview pool. When there is only one woman or underrepresented minority, s/he is far less likely to succeed than women or minorities who are compared to a diverse pool of candidates, probably because of the heightened salience of his or her race or gender.

- Rank candidates separately on several different criteria, rather than using a single aggregate ranking list. This helps mitigate the tendency for "halo" effects that result from reliance on overall impressions rather than evidence-based judgments of particular criteria.

- Consider re-opening or intensifying the search if the pool of applicants does not include female or minority candidates who will be seriously considered by the search committee.

Using Active Recruiting Practices

- Advertise the position for at least thirty days before the application deadline.

- Use electronic job-posting services targeted at diverse groups such as minority and women's caucuses or professional networks in your discipline. (A list of several resources follows on the next page.)

- Make personal contacts with women and minorities at professional conferences and invite them to apply.

- Ask faculty and graduate students to help identify women and minority candidates.

- Contact colleagues at other institutions to seek nominations of students nearing graduation or others interested in moving laterally, making sure to request inclusion of minorities and women.

- Place announcements in websites, listservs, journals, and publications aimed specifically at underrepresented minorities and women.

- Identify suitable women and minority faculty at other institutions, particularly faculty who may currently be under-placed, and send job announcements directly to them.
- Contact relevant professional organizations for rosters listing women and minorities receiving PhDs in the field.

Using Active Recruiting Resources

Be aware that most fields have resources—listservs, email groups, etc.—that can help you identify or reach qualified women and minority candidates. Either seek these out on your own, or request assistance from advance@umich.edu in identifying them.

Recruitment Sources page at Rutgers lists several resources that can be helpful in recruiting women and minority candidates.
http://uhr.rutgers.edu/ee/recruitmentsources.htm

Faculty Diversity Office page at Case Western Reserve University provides links to many specific professional organizations and diversity resources for faculty searches.
http://www.case.edu/president/aaction/diverse.html

The WISE Directories publishes free annual listings of women and minority Ph.D. recipients, downloadable as pdf documents. http://www-s.cic.net/programs/DirectoryOfWomenInScienceAndEngineering/archive/ResourceList/WiseDir/main.asp
http://www.cic.net/Home/Students/DoctoralDirectory/Introduction.aspx

The Minority and Women Doctoral Directory "is a registry which maintains up-to-date information on employment candidates who have recently received, or are soon to receive, a Doctoral or Master's degree in their respective field from one of approximately two hundred major research universities in the United States. The current edition of the directory lists approximately 4,500 Black, Hispanic, American Indian, Asian American, and women graduate students in nearly 80 fields in the sciences, engineering, the social sciences and the humanities." Directories are available for purchase.
www.mwdd.com

National Science Foundation Survey of Earned Doctorates is published yearly. While it does not list individual doctorate recipients, it is a good resource for determining how big the pool of new women and minority scholars will be in various fields.
www.nsf.gov/statistics/srvydoctorates/

Ford Foundation Fellows is an on-line directory of minority Ph.D.s in all fields, administered by the National Research Council (NRC). The directory contains information on Ford Foundation Postdoctoral fellowship recipients awarded since 1980 and Ford Foundation Predoctoral and Dissertation fellowship recipients awarded since 1986. This database does

not include Ford Fellows whose fellowships were administered by an institution or agency other than the NRC.
http://nrc58.nas.edu/FordFellowDirect/Main/Main.aspx

Mellon Minority Undergraduate Fellowship Program provides an on-line list of minority Ph.D.s and their dissertation, book and article titles in all fields.
http://www.mmuf.org/

The Faculty for the Future Project is administered by WEPAN (The Women in Engineering Program and Advocates Network), and offers a free forum for students to post resumes and search for positions and for employers to post positions and search for candidates. The website focuses on linking women and underrepresented minority candidates from engineering, science, and business with faculty and research positions at universities.
http://www.engr.psu.edu/fff/

IMDiversity.com is dedicated to providing career and self-development information to all minorities, specifically African Americans, Asian Americans, Hispanic Americans, Native Americans and women. It maintains a large database of available jobs, candidate resumes and information on workplace diversity.
http://www.imdiversity.com/

Nemnet is a national minority recruitment firm committed to helping schools and organizations in the identification and recruitment of minority candidates. Since 1994 it has worked with over 200 schools, colleges and universities and organizations. It posts academic jobs on its web site and gathers vitas from students and professionals of color.
http://www.nemnet.com

HBCU Connect.com Career Center is a job posting and recruitment site specifically for students and alumni of historically black colleges and universities.
http://jobs.hbcuconnect.com/

Society of Women Engineers maintains an online career fair.
www.swe.org

Association for Women in Science maintains a job listings page.
http://societyofwomenengineers.swe.org/

American Physical Society Education and Outreach department maintains a roster of women and minorities in physics. It contains the names and qualifications of over 3100 women and 900 minority physicists. The Roster serves as the mailing list for The Gazette, the newsletter of the APS Committee on the Status of Women in Physics (CSWP), and is widely used by prospective employers to identify women and minority physicists for job openings.
http://www.aps.org/programs/roster/index.cfm

American Indian Science & Engineering Society maintains a job listings page (and a resume database available to Career Fair exhibitors).
http://www.aises.org

American Indian Graduate Center hosts a professional organization, fellowship and post-doctoral listings, and a magazine in which job postings can be advertised.
http://www.aigcs.org

National Society of Black Engineers seeks increase the number of minority students studying engineering at both the undergraduate and graduate levels. It encourages members to seek advanced degrees in engineering or related fields and to obtain professional engineering registrations.
http://www.nsbe.org

Society of Hispanic Professional Engineers is a leading social-technical organization whose primary function is to enhance and achieve the potential of Hispanics in engineering, math and science.
http://www.shpe.org

Creating the Short List

As you begin to evaluate applicants and candidates, be aware of the kinds of evaluation biases that psychological research has identified in both women's and men's judgments of job candidates. Read Virginia Valian's book *Why So Slow?* (or some key chapters), or view her videotaped lecture summarizing this research [*http://mitworld.mit.edu/video/80*], and discuss it as a group. ADVANCE Program staff will be happy to help you obtain this material.

The most important general point about the process of creating the short list is to build in several checkpoints at which you make a considered decision about whether you are satisfied with the pool of candidates you have generated.

- Get consensus on the multiple criteria that will be used to choose candidates for interviews. Notice that different criteria may produce different top candidates. Be sure to consider all criteria that are pertinent to the department's goals (e.g., experience working with diverse students might be one). In addition, discuss the relative weighting of the different criteria, and the likelihood that no or few candidates will rate high on all of them.

- Develop a "medium" list from which to generate your short list. Are there women or minority candidates on it? If not, consider intensifying the search before moving on to a short list. Consider contacting STRIDE for advice or help.

- Consider creating separate short lists ranking people on different criteria, such as teaching, research potential, collaborative potential, and mentoring capacity. Develop your final shortlist by taking the top candidates across different criteria. Evaluate this step before finalizing the list; consider whether evaluation bias may still be affecting your choices.

- Alternatively, review the top female and/or minority candidates in your pool. Consider whether your short list should be revised because the committee's judgments were influenced by evaluation bias (the tendency to underestimate women and underrepresented minority members' qualifications and overestimate those of white males).

- Evaluation bias is minimized if you interview more than one woman and/or under-represented minority candidate. As noted earlier, research indicates that interviewers evaluate women and underrepresented minorities more fairly when there is more than one woman in the interview pool. When there is only one woman or underrepresented minority,

s/he is far less likely to succeed than women or minorities who are compared to a diverse pool of candidates, probably because of the heightened salience of his or her race or gender.

V. Handling Campus Visits

The campus visit is an important opportunity for the department to communicate three messages:

1. You are seriously interested in the candidate's scholarly credentials and work;
2. Michigan is a good place to come because it is intellectually lively, and committed to diversity in the faculty, staff and student body;
3. Michigan is a good place to come because it has a variety of humane, family-friendly policies in place.

How these messages are communicated can make a critical difference in recruiting women to departments in which they will be vastly outnumbered by male colleagues.

- Make it clear that you are interested in the candidate's scholarship and skills, rather than his or her demographic characteristics. It is generally not helpful to make a point with candidates that the department is eager to hire women and minorities.
- Consider how the department will represent the university as a whole as a place in which women and minority faculty can thrive.
- Distribute information about "family-friendly" policies (dual career, maternity leave, modified duties, etc.) to *all* job candidates regardless of gender, partner or parent status, and race or ethnicity.
- Consider how the department will represent *itself* as a place in which women and minority faculty can thrive. This may be difficult for departments that currently have few or no women and minority faculty members. Some things that may make the department more attractive to women and under-represented minorities are:
    - Clear and public policies and procedures for evaluation and promotion
    - Mentoring resources for junior faculty in general and female faculty in particular
    - Development of some practices in evaluation and annual reporting that value mentoring of women and minority faculty and students
- Schedule interviews and events with consistency in achieving outcomes, recognizing that different means may be required. For example, white male candidates may automatically be meeting with white male faculty, given the composition of your department. When recruiting candidates with different race and/or gender characteristics, it will be equally important for them to meet people who share important demographic characteristics, but you may need to make particular arrangements to ensure that this happens. Race-ethnicity and gender are not the only personal characteristics that may be important to consider; if you learn that a candidate is particularly concerned with the availability of a community identified with a particular nationality, religion, family status, sexual identity or other characteristic, take steps to help them meet with appropriate members of that community. One option is to create opportunities for the candidate to meet with faculty members, including members of STRIDE, who can provide relevant information to candidates.

- Give the candidate a chance to interact with the department's faculty in multiple venues. Formal talks may not reveal every candidate's strengths. Consider including Q + A sessions, "chalk talks," and other less formal interactions.

- Be sure to offer information and access to faculty who might represent opportunities for interdisciplinary collaboration.

- Avoid leaving candidates alone with faculty who may be hostile to hiring women and underrepresented minorities. If a candidate is confronted with racist, sexist or homophobic remarks, take positive and assertive steps to defuse the situation. Be sure there is a practice in place in the department for dealing with the expression of racist, sexist or homophobic attitudes, and that the candidate is made aware of it, if the situation arises.

- Be sure to gather equivalent information from all candidates, so you will be able to evaluate them all in terms of the same criteria. This does not require use of uniform questions with all candidates, but does require care in obtaining comparable information.

- Introduce women and minority members of the department to all candidates, not just women and minorities. Moreover, if women and minority faculty members are expected to play an especially active role in recruiting new faculty, be sure to recognize this additional service burden in their overall service load.

- Focus on the candidate's ability to perform the essential functions of the job and avoid making assumptions based on perceived race, ethnic background, religion, marital or familial status, age, disability, sexual orientation, or veteran status.

- Ask faculty to provide feedback about specific facets of the candidate's potential, rather than just requesting generic feedback. Studies show that when people focus on particular issues of performance, they are much less likely to rely on implicit biases. A sample evaluation form follows; it can be modified to represent the key criteria for your search.

Candidate Evaluation Sheet

The following offers a method for department faculty to provide evaluations of job candidates. It is meant to be a template for departments that they can modify as necessary for their own uses. The proposed questions are designed for junior faculty candidates; however, alternate language is suggested in parenthesis for senior faculty candidates.

Candidate's Name: _____

Please indicate which of the following are true for you (check all that apply):

- □ Read candidate's CV
- □ Read candidate's scholarship
- □ Read candidate's letters of recommendation
- □ Attended candidate's job talk
- □ Met with candidate
- □ Attended lunch or dinner with candidate
- □ Other (please explain): _____

Please comment on the candidate's scholarship as reflected in the job talk:

Please comment on the candidate's teaching ability as reflected in the job talk:

Please rate the candidate on each of the following:	excellent	good	neutral	fair	poor	unable to judge
Potential for (Evidence of) scholarly impact						
Potential for (Evidence of) research productivity						
Potential for (Evidence of) research funding						
Potential for (Evidence of) collaboration						
Fit with department's priorities						
Ability to make positive contribution to department's climate						
Potential (Demonstrated ability) to attract and supervise graduate students						
Potential (Demonstrated ability) to teach and supervise undergraduates						
Potential (Demonstrated ability) to be a conscientious university community member						

Other comments?

VI. Negotiating the Offer

• The way an offer is negotiated can have huge impact not only on the immediate hiring outcome, but also on a new hire's future career. Candidates who feel that chairs conduct negotiations honestly and openly, and aim to create circumstances in which they will thrive, are more satisfied in their positions and more likely to stay at the UM than are those who feel that a department or chair has deliberately withheld information, resources, or opportunities from them. Initial equity in both the negotiated conditions and in the department's follow-through on the commitments it makes are important factors in retention as well as recruitment.

• Women and minority candidates may have received less mentoring at previous career stages than their counterparts, and may therefore be at a disadvantage in knowing what they can legitimately request in negotiations. In addition, there is some evidence that women are less inclined to negotiate for themselves than men are. To ensure equity, aim to empower the candidate to advocate on his or her own behalf, by providing all candidates with a complete list of things it would be possible for them to discuss in the course of negotiations. This list will vary by field, and should include those items that will maximize the likelihood of candidate success in that field. For some fields these might include:

- Salary
- Course release time
- Lab equipment
- Lab space
- Renovation of lab space
- Research assistant
- Clerical / administrative support
- Attractive teaching opportunity
- Travel funds
- Discretionary Funds
- Summer salary
- Moving expenses
- Assistance with partner / spouse position
- Other issues of concern to the candidate

VII. Getting Off to a Good Start

• Consider appointing an advocate or mentor to help candidates throughout the negotiation process.

• Be sure to provide clear, detailed information about mentoring practices as well as all crucial review criteria and milestones such as annual reviews, third year reviews, tenure reviews, and post-tenure promotion reviews.

• If a candidate has a partner who will need placement help, try to help arrange interviews or other opportunities for the spouse or partner as early in the hiring process as possible. See the section on Dual Careers earlier, and be familiar with University resources to support these efforts. Consult the Provost's Office for further information.

VIII. Evaluating the Search

• If the department hires a woman and/or minority candidate, consider the factors that may have enabled it to do so and keep a record of good practices and successful searches for future reference.

• If the applicant pool was not as large, as qualified, or as diverse as was anticipated, consider:

- Could the job description have been constructed in a way that would have brought in a broader pool of candidates?
- Could the department have recruited more actively?
- Were there criteria for this position that were consistently not met by women or candidates of color?

• If women and/or minority candidates were offered positions that they chose not to accept, what reasons did they offer? Consider as many factors as you can identify. Are there things that the department could do to make itself more attractive to such candidates in the future? Be sure that any analysis and insight is shared with departmental decision-makers and is part of the process of initiating future searches. If you would like someone outside your department to help with a confidential interview of the candidate(s), please contact ADVANCE Program staff for help.

Appendix 1: Reading Lists

Readings on Gender, Race, Ethnicity, and Faculty Recruitment

Babcock, L. & Laschever, S. (2003). *Women don't ask: Negotiation and the gender divide.* Princeton, New Jersey: Princeton University Press.

Women Don't Ask *shows women how to reframe their interactions and more accurately evaluate their opportunities. The book includes examining how to ask for a desired outcome in ways that feel comfortable and possible, taking into account the impact of asking on relationships. It also discusses how to recognize the ways in which our institutions, child-rearing practices, and unspoken assumptions perpetuate inequalities—inequalities that are not only fundamentally unfair but also inefficient and economically unsound.*

Bauer, C.C. & Baltes, B.B. (2002). Reducing the effects of gender stereotypes on performance evaluations. *Sex Roles, 9/10,* 465–476.

This study is one of many showing (1) that people vary in the degree to which they hold certain stereotypes and schemas (2) that having those schemas influences their evaluations of other people; and (3) that it is possible to reduce the impact of commonly-held stereotypes or schemas by relatively simple means. In this study college students with particularly negative stereotypes about women as college professors were more likely to rate accounts of specific incidents of college classroom teaching behavior negatively, if they were described as performed by a female. In the second phase of the study students' reliance on their stereotypes was successfully reduced by providing them with time and instructions to recall the specific teaching behaviors of the instructors in detail. Thus, focusing attention on specific evidence of an individual's performance eliminated the previously-demonstrated effect of gender schemas on performance ratings.

Bensimon, E.M., Ward, K., & Sanders, K. (2000). *Creating mentoring relationships and fostering collegiality.* 113–137. Bolton, MA: Anker Publishing.

This section describes the department chairs' role in developing new faculty into teachers and scholars.

Bertrand, M., & Mullainathan S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review 94*(4), 991–1013; "Employers' Replies to Racial Names." NBER Website. Thursday, August 31, 2006. http://www.nber.org/digest/sep03/w9873.html.

This is an empirical study demonstrating the impact of implicit discrimination by race, and not attributable to class.

Bertrand, M., Chugh, D., & Mullainathan, D. (2005). Implicit discrimination. *American Economic Review*, *95*(2), 94–98.

This article is a reflective discussion of how and where implicit discrimination operates. Includes useful review of the literature, and fairly extended discussion of research needed.

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

Stereotypes may influence judgment via assimilation, such that individual group members are evaluated consistently with stereotypes, or via contrast, such that targets are displaced from the overall group expectation. T\vo models of judgment—the shifting standards model and status characteristics theory—provide some insight into predicting and interpreting these apparently contradictory effects. In 2 studies involving a simulated applicant-evaluation setting, we predicted and found that participants set lower minimum-competency standards, but higher ability standards, for female than for male and for Black than for White applicants. Thus, although it may be easier for low- than high status group members to meet (low) standards, these same people must work harder to prove that their performance is ability based.

Caffrey, M. (1997, May 12). Blind auditions help women. Princeton Weekly Bulletin. Based on Goldin, C & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians*. American Economic Review, 90*, 715–741.

A change in the audition procedures of symphony orchestras—adoption of "blind" auditions with a "screen" to conceal the candidate's identity from the jury—provides a test for gender bias in hiring and advancement. Using data from actual auditions for 8 orchestras over the period when screens were introduced, the authors found that auditions with screens substantially increased the probability that women were advanced (within the orchestra) and that women were hired. These results parallel those found in many studies of the impact of blind review of journal article submissions.

Chesler, M. A. (1996). Protecting the investment: Understanding and responding to resistance. *The Diversity Factor 4*(3), 2–10.

This article discusses common barriers to successful implementation of diversity-related cultural change efforts, including both those that are intentional and unintentional. It also outlines strategies for addressing or dealing with these various forms of resistance.

Cole, J. R., & Singer, B. (1991). A theory of limited differences: Explaining the productivity puzzle in science. In H. Zuckerman, J. R. Cole, and J. T. Bruer, (Eds.), *The outer circle: Women in the scientific community.* (277–310). New York: W. W. Norton and Company.

This chapter proposes "a theory of limited differences" where even if the life events to which people are exposed have small short-term effects, over the life course these events have large cumulative effects. The authors suggest that the small disparities at every stage of a woman scientist's career combine to create a subtle yet virtually unassailable barrier to success.

Dovidio, J. F. and S. L. Gaertner (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science* 11(4): 315–319.

This study investigated differences over a 10-yr period in Whites' self-reported racial prejudice and their bias in selection decisions involving Black and White candidates for employment in a sample of 194 undergraduates. The authors examined the hypothesis,

derived from the aversive-racism framework, that although overt expressions of prejudice may decline significantly across time, subtle manifestations of bias may persist. Consistent with this hypothesis, self-reported prejudice was lower in 1998-1999 than it was in 1988–1989, and at both time periods, White participants did not discriminate against Black relative to White candidates when the candidates' qualifications were clearly strong or weak, but they did discriminate when the appropriate decision was more ambiguous. Theoretical and practical implications are considered.

Fiske, S. T. (2002). What we know about bias and intergroup conflict, the problem of the century. *Current Directions in Psychological Science 11*(4): 123–128.

This essay discusses what psychologists, after years of study, now know about intergroup bias and conflict. It is stated that most people reveal unconscious, subtle biases, which are relatively automatic, cool, indirect, ambiguous, and ambivalent. Subtle biases underlie ordinary discrimination: comfort with one's own in-group, plus exclusion and avoidance of out-groups. Such biases result from internal conflict between cultural ideals and cultural biases. On the other hand, a small minority of people, extremists, do harbor blatant biases that are more conscious, hot, direct, and unambiguous. Blatant biases underlie aggression, including hate crimes. Such biases result from perceived intergroup conflict over economics and values, in a world perceived to be hierarchical and dangerous. Reduction of both subtle and blatant bias results from education, economic opportunity, and constructive intergroup contact.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902.

This article presents results of research proceeding from the theoretical assumption that status is associated with high ratings of competence, while competition is related to low ratings of warmth. Included in the article are ratings of various ethnic and gender groups as a function of ratings of competence and warmth. These illustrate the average content of the stereotypes held about these groups in terms of the dimensions of competence and warmth, which are often key elements of evaluation.

Georgi, Howard. (2000). “Is There an Unconscious Discrimination Against Women in Science?” *APS News Online*. College Park, Maryland: American Physical Society.

This is an examination of the ways in which norms about what good scientists should be like are not neutral but masculine and work to disadvantage women.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, 89(3), 416–427.

This study investigated reactions of subjects to a woman's success in a male gender-typed job. The results showed that when women were acknowledged to have been successful, they were less liked and more personally derogated than equivalently successful men. The data also showed that being disliked can affect career outcome, both for performance evaluation and reward allocation.

Katznelson, I. (2006). When affirmative action was white. *Poverty and Race Research Action Council 15*(2).

This article proposes that many federal programs can be best understood as "affirmative action for whites" both because in some cases substantial numbers of other groups were excluded from benefiting from them, or because the primary beneficiaries were whites. It states the rationale for contemporary affirmative action as "corrective action" for these exclusionary policies and programs.

Martell, R. F. (1996). What mediates gender bias in work behavior ratings? *Sex Roles 35*(3/4): 153–169.

This paper shows that more effective work behaviors are retrospectively attributed to a fictitious male police officer than a fictitious female one—even though they are rated equivalently at first. Evidence in the study shows that this results from overvaluing male officers' performance rather than derogating females'.

McNeil, L., and M. Sher. (1999). "The dual-career-couple problem." *Physics Today*. College Park, MD: American Institute of Physics.

Women in science tend to have partners who are also scientists. The same is not true for men. Thus many more women confront the "two-body problem" when searching for jobs. McNeil and Sher give a data overview for women in physics and suggest remedies to help institutions place dual-career couples.

Mickelson, R. A. and M. L. Oliver (1991). Making the short list: black faculty candidates and the recruitment process. *The Racial Crisis in American Higher Education.* C. Kerr, State University of New York Press.

This is an examination of issues involved in recruitment of racial minorities to faculty positions, especially issues associated with the prestige of training institutions.

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research and Practice, 6*, 101–115.

This article demonstrates widely-shared schemas, particularly "implicit" or unconscious ones, about race, age and gender.

Padilla, R. V. and Chavez, R. C. (1995). Introduction. The *Leaning Ivory Tower: Latino Professors in American Universities* (pp. 1–16). New York State University of New York Press.

This book includes 12 contributions from Latino and Latina professors and academics with experience in universities throughout the United States. The introduction provides an overview.

Porter, N. & Geis, F. L. (1981). Women and nonverbal leadership cues: When seeing is not believing. In C. Mayo & N. Henley (Eds.), *Gender and nonverbal behavior*. New York: Springer Verlag.

When study participants were asked to identify the leader of the group, they reliably picked the person sitting at the head of the table whether the group was all-male, all-female, or mixed-sex with a male occupying the head; however, when the pictured group was mixed-sex and a woman was at the head of the table, both male and female observers chose a male sitting on the side of the table as the leader half of the time.

Preston, A. E. (2004). *Leaving science: Occupational exit from scientific careers.* New York: Russell Sage Foundation.

Based on data from a large national survey of nearly 1,700 people who received university degrees in the natural sciences or engineering and a subsequent in-depth follow-up survey, this book provides a comprehensive portrait of the career trajectories of men and women who have earned science degrees, and addresses the growing number of professionals leaving scientific careers. Preston presents a gendered analysis of the six factors contributing to occupational exit and the consequences of leaving science.

Sagaria, M. A. D. (2002). An exploratory model of filtering in administrative searches: Toward counter-hegemonic discourses. *The Journal of Higher Education* 73(6): 677–710.

This paper describes administrator search processes at a predominately white university in order to explore whether searches may be a cause for the limited success in diversifying administrative groups.

Smith, D. (2000). How to diversify the faculty. *Academe*, 86, no. 5. Washington, D.C.: AAUP.

This essay enumerates hiring strategies that may disadvantage minority candidates or that might level the playing field.

Sommers, S. (2006). On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations. *Journal of Personality and Social Psychology* 90 (4), 597–612.

This research examines the multiple effects of racial diversity on group decision making. Participants deliberated on the trial of a Black defendant as members of racially homogeneous or heterogeneous mock juries. Half of the groups were exposed to pretrial jury selection questions about racism and half were not. Deliberation analyses supported the prediction that diverse groups would exchange a wider range of information than all-White groups. This finding was not wholly attributable to the performance of Black participants, as Whites cited more case facts, made fewer errors, and were more amenable to discussion of racism when in diverse versus all-White groups. Even before discussion, Whites in diverse groups were more lenient toward the Black defendant, demonstrating that the effects of diversity do not occur solely through information exchange. The influence of jury selection questions extended previous findings that blatant racial issues at trial increase leniency toward a Black defendant.

Steele, C. M. (1997). A threat in the air: How stereotypes shape the intellectual identities and performance of women and African-Americans. *American Psychologist, 52*, 613–629.

This paper reviews empirical data to show that negative stereotypes about academic abilities of women and African Americans can hamper their achievement on standardized tests. A 'stereotype threat' is a situational threat in which members of these groups can fear being judged or treated stereotypically; for those who identify with the domain to which the stereotype is relevant, this predicament can be self-threatening and impair academic performance. Practices and policies that can reduce stereotype threats are discussed.

Steinpreis, R.E., Anders, K.A. & Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles*, 41, 7/8, 509–528.

The authors of this study submitted the same c.v. for consideration by academic psychologists, sometimes with a man's name at the top, sometimes with a woman's. In one comparison, applicants for an entry-level faculty position were evaluated. Both men and women were more likely to hire the "male" candidate than the "female" candidate, and rated his qualifications as higher, despite identical credentials. In contrast, men and women were equally likely to recommend tenure for the "male" and "female" candidates (and rated their qualifications equally), though there were signs that they were more tentative in their conclusions about the (identical) "female" candidates for tenure.

Thompson, M. & Sekaquaptewa, D. (2002). When being different is detrimental: Solo status and the performance of women and minorities. *Analyses of Social Issues and Public Policy, 2*, 183–203.

This article spells out how the absence of "critical mass" can lead to negative performance outcomes for women and minorities. It addresses the impact on both the actor and the perceiver (evaluator).

Trix, F. & Psenka, C. (2003). Exploring the color of glass: letters of recommendation for female and male medical faculty. *Discourse & Society* 14(2): 191–220.

This study compares over 300 letters of recommendation for successful candidates for medical school faculty position. Letters written for female applicants differed systematically from those written for male applicants in terms of length, in the percentages lacking basic features, in the percentages with "doubt raising" language, and in the frequency of mention of status terms. In addition, the most common possessive phrases for female and male applicants ("her teaching" and "his research") reinforce gender schemas that emphasize women's roles as teachers and students and men's as researchers and professionals.

Turner, C.S.V.. (2002). *Diversifying the Faculty: A Guidebook for Search Committees*. Washington, D.C.: AACU.

Informed by the growing research literature on racial and ethnic diversity in the faculty, this guidebook offers specific recommendations to faculty search committees with the primary goal of helping structure and execute successful searches for faculty of color.

Valian, V. (1998). "Evaluating Women and Men." (Chapter 1 and Chapter 7.) *Why So Slow? The Advancement of Women.* Cambridge, Mass.: MIT Press.

In these chapters, Valian presents research that demonstrates that men and women who do the same things are evaluated differently, with both men and women rating women's performances lower than men's, even when they are objectively identical.

Wenneras, C. & Wold, A. (1997). "Nepotism and sexism in peer-review." *Nature*, 387, 341–343.

This Swedish study found that female applicants for postdoctoral fellowships from the Swedish Medical Research Council had to be 2.5 times more productive than their male counterparts in order to receive the same "competence" ratings from reviewers.

Wolf Wendel, L. E., S. B. Twombly, et al. (2000). "Dual-career couples: Keeping them together." *The Journal of Higher Education* 71(3): 291–321.

This paper addresses academic couples who face finding two positions that will permit both partners to live in the same geographic region, to address their professional goals, and to meet the day-today needs of running a household which, in many cases, includes caring for children or elderly parents.

Yoder, J. (2002). "2001 Division 35 Presidential Address: Context Matters: Understanding Tokenism Processes and Their Impact on Women's Work." *Psychology of Women Quarterly*, 26.

Research on tokenism processes is reviewed and coalesces around gender constructs. Reducing negative tokenism outcomes, most notably unfavorable social atmosphere and disrupted colleagueship, can be done effectively only by taking gender status and stereotyping into consideration. These findings have applied implications for women's full inclusion in male-dominated occupations.

<u>Dual career and work-family issues</u>

Boushey, H. (2005). Are Women Opting Out? Debunking the Myth. Center for Economic and Policy Research. Washington, DC, Center for Economic and Policy Research.

This analysis of the Current Population Survey's Outgoing Rotation Group data, a Bureau of Labor Statistics nationally representative survey, shows that the child penalty on labor force participation for prime-age women, aged 25 to 44, averaged -14.4 percentage points over the period from 1984 to 2004. This means that labor force participation by women in this age group with children at home averaged 14.4 percentage points less than for women without children at home. The penalty was 20.7 percentage points in 1984 and has fallen consistently over the last two decades, down to 8.2 percentage points in 2004.

Correll, S., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology* 112(5), 1297–1338.

Survey research finds that mothers suffer a substantial wage penalty, although the causal mechanism producing it remains elusive. The authors employed a laboratory experiment to

evaluate the hypothesis that status-based discrimination plays an important role and an audit study of actual employers to assess its real-world implications. In both studies, participants evaluated application materials for a pair of same-gender equally qualified job candidates who differed on parental status. The laboratory experiment found that mothers were penalized on a host of measures, including perceived competence and recommended starting salary. Men were not penalized for, and sometimes benefited from, being a parent. The audit study showed that actual employers discriminate against mothers, but not against fathers.

Goldin, C. (2006). Working it out. *The New York Times*.

Op ed article that counters the news and opinion articles claiming that women, especially graduates of top-tier universities and professional schools, are "opting out" in record numbers and choosing home and family over careers.

Kerber, L. K. (2005). We must make the academic workplace more humane and equitable. *The Chronicle of Higher Education*, 6.

This essay is a reflection by an academic historian both on the history of the academic workplace, and the ways in which it is currently an environment that is both inhumane and particularly difficult for women faculty.

McNeil, L., & Sher, M. (1999). "The Dual-Career-Couple Problem." *Physics Today*. College Park, MD: American Institute of Physics.

Women in science tend to have partners who are also scientists. The same is not true for men. Thus many more women confront the "two-body problem" when searching for jobs. McNeil and Sher give a data overview for women in physics and suggest remedies to help institutions place dual-career couples.

Radcliffe Public Policy Center (2000). *Life's work: Generational attitudes toward work and life integration.*

This paper reports on the results of a national survey of Americans' attitudes about work and family, economic security, workplace technology, and career development. The majority of young men report that a job schedule that allows for family time is more important than money, power or prestige.

Wolf Wendel, L. E., Twombly, S.B., et al. (2000). "Dual-career couples: keeping them together." *The Journal of Higher Education* 71(3): 291–321.

This article addresses academic couples who face finding two positions that will permit both partners to live in the same geographic region, to address their professional goals, and to meet the day-today needs of running a household which, in many cases, includes caring for children or elderly parents.

Background Readings on Scientific Careers

A Study on the Status of Women Faculty in Science at MIT. (1999)**.** *The MIT Faculty Newsletter*, Vol. XI, No. 4.

This is the original MIT report that has spurred so many other studies

Gannon, F., Quirk, S., & Guest, S. (2001). Are women treated fairly in the EMBO postdoctoral fellowship scheme? *European Molecular Biology Organization Reports* 2, 8, 655–657.

This article presents the findings from an analysis of the European Molecular Biology Organization Long Term Fellowship granting scheme in order to determine if gender bias exists in the program. When the success rate is calculated for the spring and autumn session for the years 1996−2001, the female applicants were, on average, 20% less successful than the males.

General Accounting Office (1994). Peer Review: Reforms Needed to Ensure Fairness in Federal Agency Grant Selection. 138.

GAO examined grant selection in three federal agencies that use peer review: the National Institutes of Health (NXH), the National Science Foundation (NSF), and the National Endowment for the Humanities (NEH). At each agency, GAO colected administrative files on a sample of grant proposals, approximately half of which had been funded. GAO then surveyed almost 1,400 reviewers of these proposals to obtain information not available from the agencies. In addition, GAO interviewed agency officials and reviewed documents to obtain procedural and policy information. GAO also observed panel meetings at each agency.

Hopkins, Nancy, Lotte Bailyn, Lorna Gibson, and Evelynn Hammonds. (2002). *An Overview of Reports from the Schools of Architecture and Planning; Engineering; Humanities, Arts, and Social Sciences; and the Sloan School of Management*. Massachusetts Institute of Technology.

The overview of MIT's more recent study of all of its schools.

Etzkowitz, H., C. Kemelgor, and B. Uzzi. (2000). "The 'Kula Ring' of scientific success." Athena unbound: The advancement of women in science and technology. Cambridge: Cambridge University Press.

This chapter and book explore the ways in which the lack of critical mass for women in science disadvantages them when it comes to the kinds of networking that promotes collaboration and general flow of information needed to foster the best possible research.

Kulis, S., Chong, Y., & Shaw, H. (1999). Discriminatory organizational contexts and black scientists on postsecondary faculties. *Review in Higher Education, 40*(2), 115–148.

This article examines the role of various kinds of institutional discrimination in producing the underrepresentation of black faculty.

Long, J. Scott, ed. (2001). Executive summary. From scarcity to visibility: Gender differences in the careers of doctoral scientists and engineers. 1–8. Washington, D.C.: National Academy Press.

This excerpt provides an overview of differences in the science careers of men and women.

Mervis, J. (2005). A glass ceiling for Asian scientists? *Science, 310,* 606–607.

This article documents the low rate of Asian and Asian American scientists at higher and leadership levels even in fields where they are relatively numerous at lower ranks.

Nelson, D. J., & Rogers, D. C. (2004). A national analysis of diversity in science and engineering faculties at research universities.

This report looks at the representation of women and minorities in the 'top 50' departments of science and engineering disciplines in research universities, as ranked by the National Science Foundation according to research funds expended. The report is based on survey data obtained from these departments and covers the years 1993 to 2002. The analysis examines degree attainment (BS and PhD) and representation on the faculty in the corresponding disciplines. The data demonstrate that while the representation of women attaining a PhD in science and engineering has significantly increased in this period, the corresponding faculties remain overwhelmingly dominated by white men.

Appendix 2: Active Recruiting Resources

Be aware that most fields have resources—listservs, email groups, etc.—that can help you identify or reach qualified women and minority candidates. Either seek these out on your own, or request assistance from advance@umich.edu in identifying them. Some fairly broad listings are included here.

"Guidelines for Recruiting a Diverse Workforce." Penn State University. Available online: www.psu.edu/dept/aaoffice/pdf/guidelines.pdf

"Faculty Recruitment Toolkit." (2001). University of Washington. Available online: http://www.engr.washington.edu/advance/resources/FacultyRecruitmentToolkit_20080205.pdf

"Recruitment and Selection of Faculty and Academic Professional and Administrative Employees
Appendix A: Recruiting a Diverse Qualified Pool of Applicants" University of Minnesota. Available online:
http://policy.umn.edu/groups/hr/documents/appendix/recruitfacpa_appa.pdf

"Massachusetts Institute of Technology Faculty Search Committee Handbook." (2002). Available online as pdf document:
http://web.mit.edu/faculty/reports/FacultySearch.pdf

"Search Committee Toolkit." University of California at Los Angeles. Available Online:
http://faculty.diversity.ucla.edu/search/searchtoolkit/docs/SearchToolkit071008.pdf

"Faculty Search Committee Guidelines." Case Western Reserve University. Available Online as pdf document:
http://www.case.edu/president/aaction/Faculty%20Search%20Guide.pdf

"Recruitment and Retention: Guidelines for Chairs." (updated 2007). Hunter College, CUNY.
Available online as pdf document:
http://www.hunter.cuny.edu/genderequity/equityMaterials/Jan2007/recruitretain.107.pdf

"Leap Recruiting Faculty Brochure." University of Colorado, Boulder. Available online as pdf document:
http://www.colorado.edu/facultyaffairs/leap/downloads/leap_recruiting.pdf

The WISE Directories publishes free annual listings of women and minority Ph.D. recipients, downloadable as pdf documents.
http://www-s.cic.net/programs/DirectoryOfWomenInScienceAndEngineering/archive/ResourceList/WiseDir/main.asp
http://www.cic.net/Home/Students/DoctoralDirectory/Introduction.aspx

The Minority and Women Doctoral Directory "is a registry which maintains up-to-date information on employment candidates who have recently received, or are soon to receive, a Doctoral or Master's degree in their respective field from one of approximately two

hundred major research universities in the United States. The current edition of the directory lists approximately 4,500 Black, Hispanic, American Indian, Asian American, and women graduate students in nearly 80 fields in the sciences, engineering, the social sciences and the humanities." Directories are available for purchase.
http://www.mwdd.com/

National Science Foundation Survey of Earned Doctorates is published yearly. While it does not list individual doctorate recipients, it is a good resource for determining how big the pool of new women and minority scholars will be in various fields.
www.nsf.gov/statistics/srvydoctorates/

Ford Foundation Fellows is an on-line directory of minority Ph.D.s in all fields, administered by the National Research Council (NRC). The directory contains information on Ford Foundation Postdoctoral fellowship recipients awarded since 1980 and Ford Foundation Predoctoral and Dissertation fellowship recipients awarded since 1986. This database does not include Ford Fellows whose fellowships were administered by an institution or agency other than the NRC.
http://nrc58.nas.edu/FordFellowDirect/Main/Directory.aspx

Mellon Minority Undergraduate Fellowship Program provides an on-line list of minority Ph.D.s and their dissertation, book and article titles in all fields.
http://www.mmuf.org/ (select Fellows Update from the menu bar on the main page)

The Faculty for The Future Project is administered by WEPAN (The Women in Engineering Program and Advocates Network), and offers a free forum for students to post resumes and search for positions and for employers to post positions and search for candidates. The website focuses on linking women and underrepresented minority candidates from engineering, science, and business with faculty and research positions at universities.
http://www.engr.psu.edu/fff/

IMDiversity.com is dedicated to providing career and self-development information to all minorities, specifically African Americans, Asian Americans, Hispanic Americans, Native Americans and women. It maintains a large database of available jobs, candidate resumes and information on workplace diversity.
http://www.imdiversity.com/

Nemnet is a national minority recruitment firm committed to helping schools and organizations in the identification and recruitment of minority candidates. Since 1994 it has worked with over 200 schools, colleges and universities and organizations. It posts academic jobs on its web site and gathers vitas from students and professionals of color.
http://www.nemnet.com

HBCU Connect.com Career Center is a job posting and recruitment site specifically for students and alumni of historically black colleges and universities.
http://jobs.hbcuconnect.com/

Society of Women Engineers maintains an online career fair.
www.swe.org

Association for Women in Science maintains a job listings page.
www.awis.org

American Indian Science & Engineering Society maintains a job listings page (and a resume database available to Career Fair exhibitors).
http://www.aises.org

American Indian Graduate Center hosts a professional organization, fellowship and post-doctoral listings, and a magazine in which job postings can be advertised.
http://www.aigcs.org

National Society of Black Engineers
http://www.nsbe.org

Society of Hispanic Professional Engineers
http://www.shpe.org

American Physical Society Education and Outreach department maintains a roster of women and minorities in physics. It contains the names and qualifications of over 3100 women and 900 minority physicists. The Roster serves as the mailing list for The Gazette, the newsletter of the APS Committee on the Status of Women in Physics (CSWP), and is widely used by prospective employers to identify women and minority physicists for job openings.
http://www.aps.org/programs/roster/index.cfm

Recruitment Sources page at Rutgers lists several resources that can be helpful in recruiting women and minority candidates.
http://uhr.rutgers.edu/ee/recruitmentsources.htm

Faculty Diversity Office page at Case Western Reserve University provides links to many specific professional organizations and diversity resources for faculty searches.
http://www.case.edu/president/aaction/diverse.html

This material is based upon work originally supported by the National Science Foundation under Grant Number SBE-0123571. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

cepr

Are Women Opting Out? Debunking the Myth

BY HEATHER BOUSHEY

Executive Summary

A front page article in the *New York Times* (Story 2005) recently reported that women at Yale University are planning to quit their jobs when they enter their thirties and have children. Both *Time* (Wallis 2004) and the *New York Times Magazine* (Belkin 2003) have recently featured cover articles arguing that some mothers—especially older, highly educated new mothers—are increasingly likely to stop working when they have kids. These articles reflect the popular notion that women are increasingly "opting out" of employment when they have children. The basic argument is that mothers are choosing to stay at home in greater numbers due to the stresses of living in two-earner families or making it in the labor market.

Such news coverage may lead people to believe that there is a growing trend toward this sort of "opt-out" by highly educated mothers. However, economic data provides no evidence to support these anecdotal accounts. In 2004, the latest year for which a full year of data is available, the impact of having children in the home on women's labor force participation (the "child penalty") fell compared to prior years.

This analysis of the Current Population Survey's Outgoing Rotation Group data, a Bureau of Labor Statistics nationally representative survey, shows that the child penalty on labor force participation for prime-age women, aged 25 to 44, averaged -14.4 percentage points over the period from 1984 to 2004. This means that labor force participation by women in this age group with children at home averaged 14.4 percentage points less than for women without children at home. The penalty was 20.7 percentage points in 1984 and has fallen consistently over the last two decades, down to 8.2 percentage points in 2004.

In terms of the current labor market situation, the data show that, after controlling for changes in demographics and the labor market, the negative effects of children on women's labor supply fell between 2000 and 2004. In 2004, prime-age women with children at home were 8.2 percentage points less likely to be in the labor force than were women without children, but this was down from a 9.9 percentage-point penalty in 2000. The analysis finds a similar downward trend for women

Center for Economic and Policy Research
1611 Connecticut Ave, NW
Suite 400
Washington, DC 20009
tel: 202-293-5380
fax:: 202-588-1356
www.cepr.net

Heather Boushey is an Economist at the Center for Economic and Policy Research. John Schmitt and Todd Tucker provided valuable insight and comments on this paper and Ben Zipperer provided research assistance.

with small children, across age groups, and across educational attainment levels, except for women in their thirties with professional or advanced degrees, for whom there is no statistically significant change over time.

The data stand in opposition to the media frenzy on this topic. In spite of the personal anecdotes highlighted in various news stories, women are not increasingly dropping out of the labor force because of their kids. The main reasons for declining labor force participation rates among women over the last four years appears to be the weakness of the labor market.

The recession of the early 2000s led to sustained job losses for all women–those with and without children at home–and by early 2005 the labor market had only just returned to its 2000 employment level, almost exactly four years after the recession began. During this recession, women experienced their largest employment losses in decades and once this is controlled for, the presence of children at home plays a smaller role in women's labor force participation than it did in previous years, going back to 1984.

The effect of children on women's labor force participation

This analysis addresses whether or not a woman with a child in the home is any less likely to be in the labor force today than she was at earlier points in the last two decades simply because there was a child in her household. To examine this question, this analysis uses data from the Center for Economic and Policy Research (CEPR) Outgoing Rotation Group (ORG) Extracts for years from 1984 to 2004.[1] The analysis examines the effects of children on the labor force participation rates of prime-age women, aged 25 to 44, regardless of their biological relationship to the child, including un-adopted stepchildren and foster children, along with biological children. Thus, the paper will refer to "mothers" as "women with children at home."

The focus is on women's labor force participation rates (LFPR), which is the share of a population either at work (employed) or actively searching for work (unemployed). To the extent that journalists and other commentators on the child penalty have used nationally representative data to support their claims, they have used "raw" (or "unadjusted") LFPR rates, which compare the LFPR rates over time without controlling for the changing demographic characteristics of women or the cyclical nature of the overall labor market. For example, over the past two decades, the share of women who identify themselves as Hispanic, not black, has more than doubled. The rise in population of Hispanic women could pull down mothers' LFPRs because Hispanic mothers typically are much more likely than other mothers to be full-time homemakers. Thus, this demographic shift alone could, if the magnitudes were large enough, explain changes in women's LFPRs over time. This analysis seeks to isolate the effects of children on women's LFPRs, independent of any changes in the composition of the population of women or cyclical changes in the labor market. Therefore, the focus will be on LFPRs that "control" for factors, such as race and ethnicity, age, education, and the business cycle.

---

[1] The CEPR ORG Extracts are publicly available and fully documented at www.ceprdata.org. These Extracts are from the Bureau of Labor Statistics Current Population Survey.

Several demographic and labor market changes over the past two decades may have influenced women's and mother's LFPRs. Immigration, changes in educational attainment among women, increased divorce rates, an aging labor force, and increases in mothers' age at first birth are all factors that may affect a woman's decision to work. Table 1 shows that over the two decades from 1984 to 2004, the average age among women in the 25 to 44 age range increased about 1.5 years, while the share of women in that same age range with children at home decreased (down about 0.5 percentage points for women with children under six; and over 3 percentage points for women with children under 18). Overall, women in all age ranges are also now better educated, though few (8.6 percent) have a graduate degree.

Changes in the composition of the population of women took place against normal cyclical changes in the U.S. economy. Between 2000 and 2004, for example, the labor market went from a peak through a prolonged period of slow economic recovery. As of October 2005, the employment rate—the share of the total population at work—remained 1.8 percentage points below its cyclical peak in 2000. The lackluster labor market and demographic changes have both affected women's LFPRs, leading to the illusion in the raw data that children have caused women to drop out of the labor market.

Table 1. Characteristics of women aged 25 to 44

	Year	
	1984	2004
Average age	34.0	35.5
	Share (%)	
Has children under age six	30.2	29.7
Has any children	66.9	63.1
Married	73.8	62.6
Race/ethnicity		
White	78.0	64.8
Black	12.0	13.4
Hispanic	6.7	15.1
Other	3.3	6.8
Educational attainment		
Less than high-school	14.7	9.8
High-school	40.5	30.1
Some college	23.3	29.3
College degree	15.8	22.2
Graduate degree	5.6	8.6

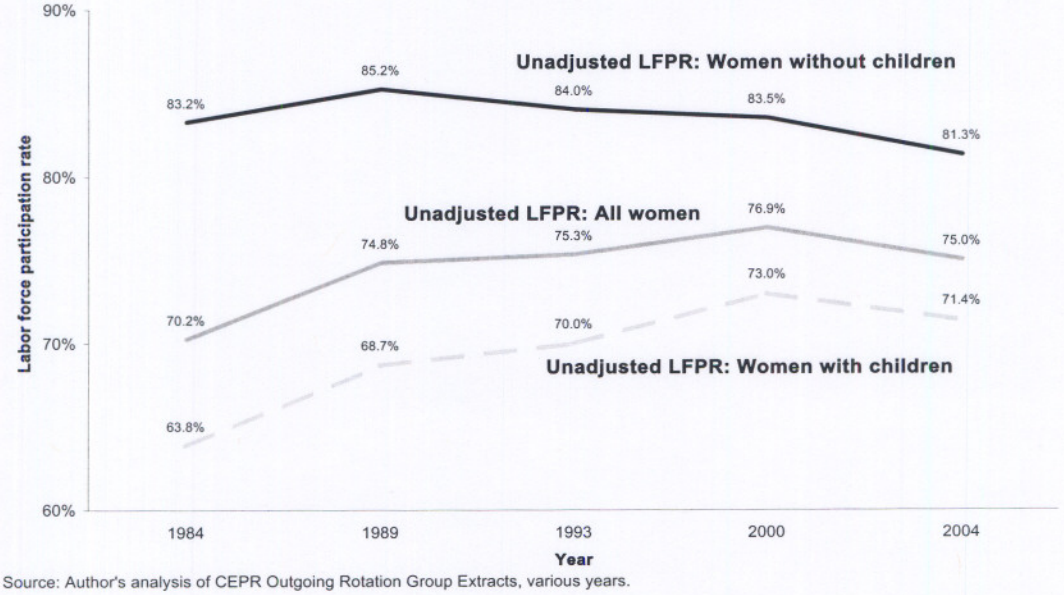
Source: Author's analysis of CEPR Outgoing Rotation Group Extracts, various years.

The tables and figures that follow show the raw LFPR, not adjusted for demographics or other factors, alongside estimates of the specific effects of children on women's LFPR, controlling for characteristics of the mother and the business cycle. The probability that a prime-age woman is in the labor force if she has a child is estimated using a standard statistical procedure (logit model), which controls for the effects of age, race and ethnicity, educational attainment, marital status, whether there is a prime-age male in the household in the labor force, and the year (to capture the effects of the business cycle on all women, those with and without children at home).[2] This model provides a separate estimate for each year of the "marginal effect" of having a child on women's likelihood of being in the labor force. The paper will refer to this effect, which is the percentage point change in the probability of being in the labor force due specifically to having a child in each year, as the "child penalty."

Women's labor force participation rates have not fallen due to the presence of children at home

Figure 1 shows the unadjusted LFPRs for all women, and women with and without children at home. All three groups experienced a sharp drop in the unadjusted LFPR between 2000 and 2004. The percentage-point decline was actually slightly larger for women without children—2.2 percentage points—than it was for women with children—1.6 percentage points. The drop in the LFPR for women with children has led to claims that mothers are opting out of employment. Figure 1, however, does not isolate the effects of children on women's LFPR; it only says that women who have children are less likely to be in the labor force.

Figure 1.
Women's unadjusted labor force participation rate (LFPR), aged 25 to 44, by presence of children at home, 1984 to 2004



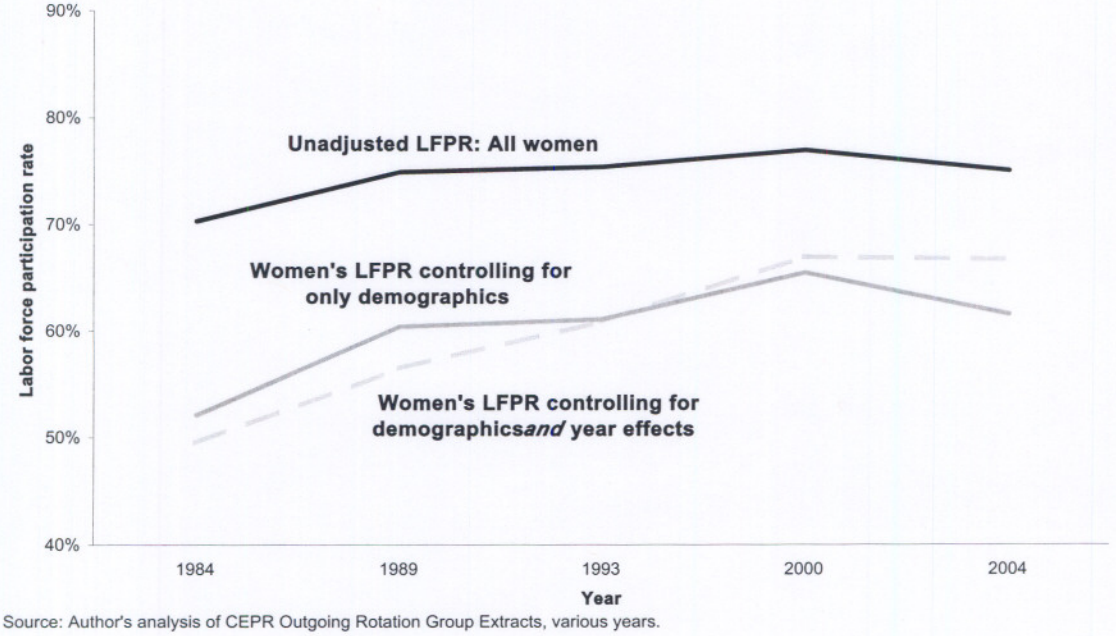
Source: Author's analysis of CEPR Outgoing Rotation Group Extracts, various years.

[2] Full model results are available from the author upon request.

Figure 2a shows the unadjusted LFPR for all women alongside estimates of women's LFPR, including the "child penalty" (calculated using the logit model that controls for demographic characteristics and the business cycle). The model that estimates the child penalty by controlling for demographics only (age, education, marital status, and a man working in the household) continues to show a decline in LFPR between 2000 and 2004. However, the model that estimates the child penalty including demographics and the year (as a proxy for the business cycle) shows no increase in women opting out of the labor force due specifically to the presence of children in the home.

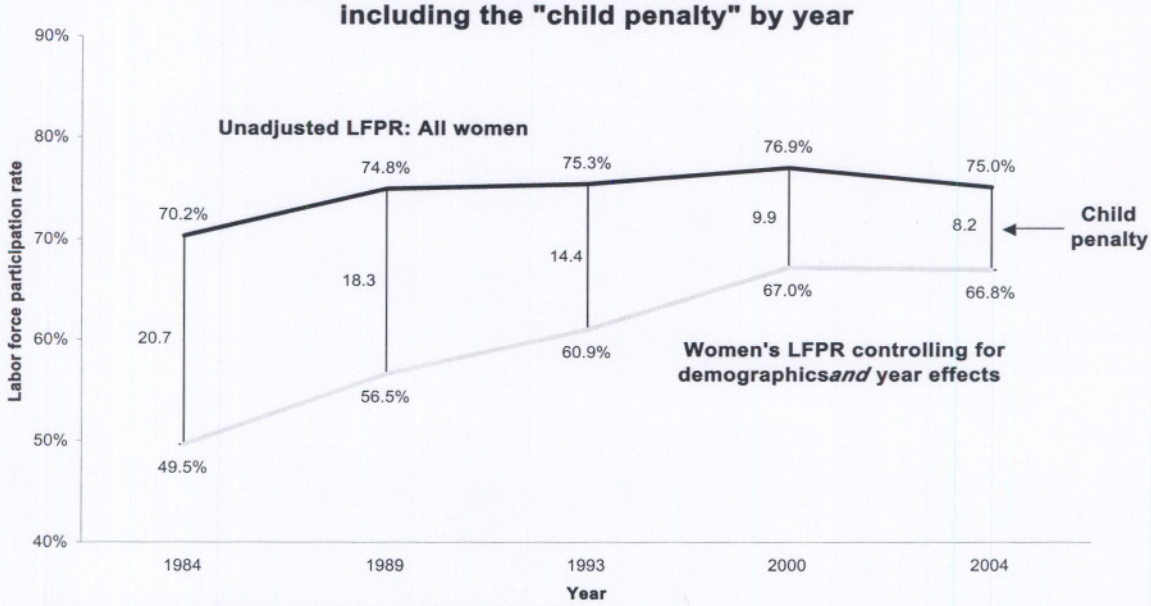
Figure 2b presents the same information as 2a, but highlights the statistically adjusted effects of children on women's LFPR. Between 2000 and 2004, after controlling for changes in the composition of the population of women and the general decline in the labor market, the child penalty actually fell from 9.9 to 8.2 percentage points (statistically significant at the one percent level). All women have seen their labor supply fall between 2000 and 2004, which is something that economists need to explain. However, the claim that this decline is explained by the decision of women to stay home with their children is simply not true. Children had no more impact—if anything a smaller impact—on women's decisions to join the labor force in 2004 than they did at any earlier point in the preceding 20 years.

Figure 2a.
Labor force participation rates (LFPR), adjusted and unadjusted, women, aged 25 to 44, with any children at home, 1984 to 2004



Source: Author's analysis of CEPR Outgoing Rotation Group Extracts, various years.

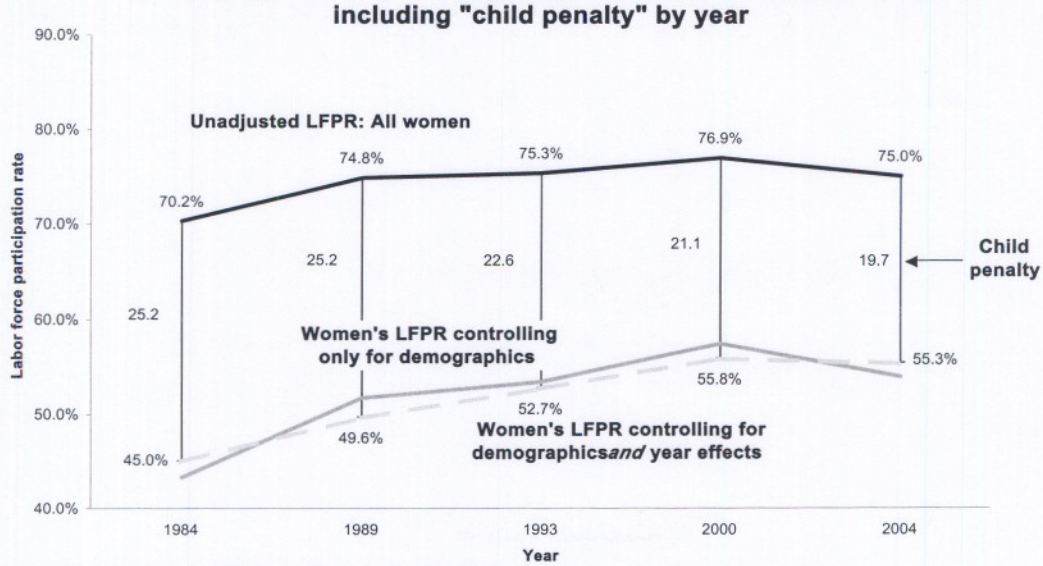
Figure 2b.
Labor force participation rates (LFPR), adjusted and unadjusted, women, aged 25 to 44, with any children at home, 1984 to 2004, including the "child penalty" by year



Source: Author's analysis of CEPR Outgoing Rotation Group Extracts, various years.

The trends are similar for women with young children (under age six) at home (see Figure 3). As would be expected, fewer women with young children are in the labor force than is the case for women with any children under age 18. The effect of young children on women's participation decision, however, was smaller in 2004 than in 2000 (statistically significant at the one percent level), or any other period examined since 1984.

Figure 3.
Labor force participation rates (LFPR), adjusted and unadjusted, women, aged 25 to 44, with young children at home, 1984 to 2004, including "child penalty" by year



Source: Author's analysis of CEPR Outgoing Rotation Group Extracts, various years.

Women's labor force participation rates have fallen due to the early 2000s recession

Because the raw data do not take into account demographic and critical business cycle effects, the raw data mask the fact that women's LFPRs have been pushed downwards by the prolonged period of slow labor market recovery. The recession of the early 2000s was harder on women than the recessions of the 1980s or 1990s, and in particular, harder on younger women (Boushey, Rosnick, and Baker 2005). The effect of higher job losses on women's labor force participation can be seen in Table 2, which shows the relative effects of the business cycle on prime-age and younger women's labor force participation rates, controlling for women's demographic characteristics. Compared to 1984, prime-age women were 3.2 percentage points more likely to be employed in 2000, but only 1.1 percentage points more likely to be employed in 2004. Thus, women were overall less likely to be in the labor force in 2004, compared to 2000 (3.2 percent versus 1.1 percent).

Table 2. Year effects on women's labor force participation rates
(Standard errors in parentheses)

	Women's labor force participation, controlling for demographics (percentage point)		
	Prime-age women (aged 25 to 44)	Women aged 25 to 32	Women aged 33 to 39
Year effects, relative to 1984			
1989	3.2 (0.3)*	2.3 (0.5)*	2.9 (0.5)*
1993	2.5 (0.3)*	0.6 (0.5)	3.1 (0.5)*
2000	3.6 (0.4)*	3.0 (0.6)*	3.2 (0.6)*
2004	1.1 (0.4)*	-0.0 (0.6)	1.0 (0.6)^

Source: Author's analysis of CEPR Outgoing Rotation Group Extracts, various years.
* significant at the 1 percent level; ^ significant at the 10 percent level.
Notes: Marginal effects are derived from a logit regression on women's labor force participation. Other independent variables in the model are dummies for age, educational attainment, race/ethnicity, and whether there is a prime-age male in the labor force in the household. Full results available from the author.

For younger women, the early 2000s recession had a stronger dampening effect on labor force participation. Compared to 1984, in 2004, women aged 25 to 32 were no more or less likely to be in the labor force. In 2000, women points more likely to be in the labor force in 2000 compared to 1984, but only 1.0 percentage points more likely to be in the labor force in 2004.

Higher job losses in the recession of the early 2000s have had the effect of making it appear that women—and especially women with children—are opting out of employment. Table 3 shows the effects of having children in the home on women's labor supply, with and without controls for the year. Column 1 shows that the overall effect of having a child under age 18 in the home on prime-age women is to reduce the probability of being in the labor force by 14.4 percentage points. Column 2 looks at the child penalty by year, including only demographics, but not a separate business cycle

effect. Here, we see that the child penalty decreases from 1984 to 2000, from 18.2 percentage points down to 11.5 percentage points. Between 2000 and 2004, however, the child penalty rises back up to 13.5 percent. Column 3 shows that there is no such decline once the model includes a year effect, to control for the business cycle, as well as cultural and other changes affecting all women (not just mothers) that may have occurred over time. The year effects shown in Column 3 are the effects of the business cycle of women's LFPR: in 2004, compared to 1984, women were 7.4 percentage points less likely to work because of the weak labor market. This dampening effect on women's LFPRs is nearly as large as the 8.2 percentage point child penalty in 2004.

Table 3. Child penalty on women's labor force participation rate
(Standard errors in parentheses)

	Prime-age women (aged 25 to 44)		
	(1)	(2)	(3)
Has child at home	-14.4		
	(0.3)*		
Has child at home, by year			
1984		-18.2	-20.7
		(0.4)*	(0.6)*
1989		-14.5	-18.3
		(0.4)*	(0.6)*
1993		-14.3	-14.4
		(0.4)*	(0.6)*
2000		-11.5	-9.9
		(0.4)*	(0.6)*
2004		-13.5	-8.2
		(0.4)*	(0.6)*
Year effects, relative to 1984			
1989			1.2
			(0.6)^
1993			-2.1
			(0.6)*
2000			-3.8
			(0.7)*
2004			-7.4
			(0.7)*

Source: Author's analysis of CEPR Outgoing Rotation Groups Extracts, various years.
* significant at the 1 percent level; ^significant at the 10 percent level.
See notes to Table 2.

Labor participation rates for highly-educated women in their thirties are, for the most part, unchanged

Much of the focus of the discussion on women's labor force participation has been on one specific group of mothers: highly educated, older, first-time mothers. Again, there is no evidence of this trend in the national statistics. Children have a very small effect on highly educated women's LFPR, which has continued to rise up through 2004.

Table 4 shows the child penalty for women in their late twenties and early thirties, and women aged 33 to 39. For women in their late twenties and early thirties, the model that includes year effects shows that between 1984 and 2004, the child penalty fell by more than half, from 27.3 down to 11.9 percentage points. For women in their thirties, over this time period the child penalty has fallen by nearly two-thirds, from 18.2 down to 7.0 percentage points. The child penalty for thirty-something women was essentially unchanged between 2000 and 2004. Overall, older women face a smaller child penalty compared to younger women. This is true even if the woman has young children at home (results not shown).

Table 4. Child penalty on women's labor force participation rate, by woman's age
(Standard errors in parentheses)

	Women aged 25 to 32			Women aged 33 to 39		
	(1)	(2)	(3)	(4)	(5)	(6)
Has child at home	-20.4 (0.5)*			-12.2 (0.5)*		
Has child at home, by year						
1984		-24.1 (0.6)*	-27.3 (0.9)*		-15.3 (0.6)*	-18.2 (1.1)*
1989		-20.6 (0.6)*	-24.0 (0.9)*		-11.9 (0.7)*	-14.6 (1.1)*
1993		-20.9 (0.6)*	-18.8 (0.9)*		-11.5 (0.6)*	-12.6 (1.0)*
2000		-17.0 (0.7)*	-16.1 (1.0)*		-10.1 (0.7)*	-7.9 (1.0)*
2004		-18.7 (0.7)*	-11.9 (1.0)*		-12.2 (0.7)*	-7.0 (1.0)*
Year effects, relative to 1984						
1989			0.2 (0.9)			-0.2 (1.2)
1993			-4.5 (0.9)*			-1.5 (1.1)
2000			-3.5 (1.0)*			-4.6 (1.3)*
2004			-8.9 (1.0)*			-7.4 (1.3)*

Source: Author's analysis of CEPR Outgoing Rotation Groups Extracts, various years.
* significant at the 1 percent level.
See notes to Table 2.

What is striking about Table 4 is that there are strong cyclical effects on women's LFPRs. In 2004, the weak labor market relative to 1984 led to an 8.9 percentage point reduction in younger women's LFPRs, almost as large as the child penalty of 11.9 percentage points in 2004. For women in their thirties, the weak labor market of 2004 led to a 7.4 percentage point reduction in LFPR compared to 1984, which was larger than the 2004 child penalty of 7.0 percentage points. In 2004, women's LFPRs were pushed down because of poor performance in the labor market, not because of an increased desire to stay home with children.

The media, however, has focused not simply on relatively older mothers, but on highly educated, thirty-something mothers. This is a tiny share of all women with children: 96.8 percent of women aged 25 to 44 with children are *not* women in their mid- to late-thirties who have advanced degrees. While it is true that this group of women has increased among prime-age mothers—from 2.5 to 3.2 percent—the real story of how "American mothers" are balancing work and family cannot be found in the trends of such a small and extremely advantaged group. This group of mothers is advantaged compared to other prime-age mothers in terms of not only educational attainment and earnings potential, but they are more likely to be married (91.2 percent versus 78.3 percent of all mothers aged 25 to 44) and are highly likely to have a spouse who also has very high earnings potential. Further, they are more likely to be in the kinds of jobs that provide the benefits and workplace flexibility that makes work/family balance not entirely an oxymoron (Boushey 2005).

Most thirty-something mothers work. Not only are highly-educated, thirty-something women with children at home a relatively small share of the population, but, compared to other educational groups, they are also more likely to be in the labor force if they have children and their child penalty is smaller than for other educational groups—so small as to be statistically insignificant in most years. However, it is also the case that the majority of highly-educated, thirty-something women who are not at work have children at home. In short, the overwhelming majority of thirty-something women with advanced degrees do not opt out if they have kids, but if they do opt out, they have kids. This is less the case for other women, making this group truly exceptional.

Tables 5 and 6 explore the LFPRs of women by age and educational attainment, documenting just how exceptional highly-educated, thirty-something women are compared to other women with children. Column 1 of Table 5 shows that thirty-something, highly-educated women are more likely than other educational groups to be in the labor force if they have children at home—even young children. Nearly three-quarters (73.2 percent) of highly-educated women in their thirties with a young child at home are in the labor force, controlling for other demographic characteristics and cyclical effects. Nearly eight-out-of-ten women with a graduate degree who have a child under age 18 at home are in the labor force, a rate higher than for all other educational groups in this age range. For example, only 60.2 percent of women with a high-school degree who have a child at home are in the labor force, controlling for demographics and the business cycle.

Even though adjusted LFPRs for women with children are higher for highly-educated women, it is true that this group is also more likely to have children at home if they are not working. Column 3 of Table 5 shows that nearly nine-out-of-ten (87.0 percent) highly-educated women in their thirties who are out of the labor force have a child at home; over three-quarters (75.7 percent) have a young child

at home. This is higher than other among other educational groups: among women in their thirties with a high-school degree, a third (36.7 percent) of those not working have a young child at home. Thus, the group of highly-educated women is so attached to the labor force that, for most of them, having a child may be the only reason not to work.

If working, however, highly-educated thirty-something women are less likely than other women to have children at home. Among those in the labor force, three-quarters (76.2 percent) of women without a high-school degree have a child at home, while only six-out-of ten (57.7 percent) women with a graduate degree have a child at home. Highly-educated women are more likely to have a small child at home, compared to other educational groups, because they wait longer to have children, compared to other women. Four-out-of-ten (40.3 percent) highly-educated women have a child under age six at home, compared to only one-in-five (22.0 percent) of those with a high-school degree.

Table 5. Labor force participation, by educational attainment and children

	(1)	(2)	(3)
		Share with children	
	Adjusted LFPR	In labor force	Not in labor force
	2004	2004	2004
Any child			
Less than high-school	28.8%	76.2%	76.3%
High-school	60.2	73.6	75.9
Some college	72.5	72.3	82.4
College degree	71.3	63.1	83.5
Graduate degree	77.0	57.7	87.0
Child under age six			
Less than high-school	18.2	20.3	37.1
High-school	55.6	22.0	36.7
Some college	66.6	26.5	49.2
College degree	66.6	37.1	65.6
Graduate degree	73.2	40.3	75.7

Source: Author's analysis of CEPR Outgoing Rotation Group Extracts, various years.

Higher LFPRs among highly-educated women with children occurs alongside far smaller child penalties, compared to other educational groups. Table 6 examines the effects of children on women's LFPR for prime-age women and women in their thirties, by educational attainment. The estimates are derived from regressions that control for demographic and business cycle effects. Among prime-age women with a graduate degree, the penalty for having a child has changed little over the past two decades, and only in peak years—1989 and 2000—is the penalty statistically significant. Over the past two decades, among all prime-age women in all educational groups, there has not been any increase of a statistically significant magnitude or not in the child penalty on women's LFPR.

The trends are generally the same among women in their thirties. Among all education groups except women with graduate degrees, the child penalty has fallen each year. For women with graduate degrees, the penalty rose from 2000 to 2004, however, this change was statistically insignificant. By comparison, for women with a college degree, the drop in the child penalty from 7.9 to 3.8 percentage points was statistically significant at the one percent level. (The significance in the table notes only if the coefficient is significant overall; a separate test of statistical difference between the coefficients on presence of children in 2000 and 2004 found that it could not be ruled out that they were the same.) Overall, the child penalty for the most highly educated group of women is small, and not generally significant, whereas the child penalty is larger for less-educated women and is statistically significant within each year. Therefore, while attention focuses on whether an elite group of women are opting out of employment, in reality, the child penalty on labor force participation is smallest for this group and they have the highest LFPRs.

Table 6. Child penalty, by woman's age and educational attainment level

	Less than high-school	High-school	Some college	College degree	Graduate school
Prime-age women (aged 25 to 44)					
1984	-30.5	-16.6	-13.2	-12.6	-1.4
	(0.8)*	(0.6)*	(0.6)*	(0.7)*	(0.9)
1989	-30.7	-14.2	-11.0	-10.9	-2.2
	(0.9)*	(0.6)*	(0.6)*	(0.7)*	(1.0)#
1993	-31.0	-11.5	-7.2	-7.0	-0.6
	(0.9)*	(0.5)*	(0.5)*	(0.7)*	(0.8)
2000	-21.7	-8.0	-3.9	-7.2	-1.5
	(1.0)*	(0.6)*	(0.5)*	(0.7)*	(0.9)^
2004	-21.7	-7.2	-2.8	-4.6	-1.3
	(1.0)*	(0.5)*	(0.5)*	(0.6)*	(0.8)
Women aged 32 to 39					
1984	-25.0	-14.1	-10.7	-11.3	-1.4
	(1.5)*	(1.0)*	(1.1)*	(1.2)*	(1.3)
1989	-26.6	-10.0	-8.0	-9.3	0.1
	(1.6)*	(0.9)*	(1.0)*	(1.1)*	(1.2)
1993	-25.9	-9.6	-5.6	-6.9	-0.5
	(1.6)*	(0.9)*	(0.8)*	(1.0)*	(1.2)
2000	-15.7	-5.6	-2.2	-7.9	-0.6
	(1.5)*	(0.9)*	(0.8)*	(1.0)*	(1.2)
2004	-18.8	-5.3	-2.0	-3.8	-2.5
	(1.6)*	(0.8)*	(0.8)*	(0.9)*	(1.2)#

Source: Author's analysis of CEPR Outgoing Rotation Groups Extracts, various years.
* significant at the 1 percent level; # significant at the 5 percent level; ^ significant at the 10 percent level.
See notes to Table 2.

The most recent labor market recession has also disproportionately hurt less-educated mothers. Table 7 shows the percentage point decline in women's LFPR in 2004, relative to 1984, by educational attainment. The story here is that, controlling for demographics and the child penalty, less-educated women have a higher "business cycle penalty" on their LFPR, compared to women with advanced

degrees. In 2004, among women with any children, the business cycle penalty was 7.6 percentage points for women with a high-school degree and 3.5 percentage points for women with a college degree, while women with an advanced degree have only a 2.3 percentage point penalty. The business cycle penalty is significantly greater in 2004 than in 2000 for all educational groups except for women with advanced degrees and either young or older children and women with less than a high-school degree and any children. Thus, compared to women with advanced degrees, the business cycle penalty is not only larger in 2004, but also significantly greater than in 2000, showing just how difficult this recession has been on the majority of working women.

Table 7. Labor market effect on women's LFPR, 2004

(Percentage point change relative to 1984)

	Prime age women (aged 25 to 44)	
Educational attainment	Has any children	Has child under age six
Less than high-school	-4.4	-6.9
	(1.8)#	(13.5)^
High-school	-7.6	-9.0
	(0.9)*	(1.8)*
Some college	-4.6	-3.9
	(0.9)*	(01.6)#
College degree	-3.5	-2.2
	(0.6)*	(1.1)#
Graduate degree	-2.3	-1.2
	(1.1)#	(0.9)

Source: Author's analysis of CEPR Outgoing Rotation Groups Extracts, various years.

* significant at the 1 percent level; # significant at the 5 percent level; ^ significant at the 10 percent level.

See notes to Table 2.

Conclusion

The important trend that this paper has explored is that the lackluster performance of the labor market since 2001 is the real reason that LFPRs have been falling among women. Women's employment rates were hit exceptionally hard by this recession and they have yet to fully recover. While women had previously been more insulated from cyclical unemployment, compared to men, now they appear to be nearly as vulnerable, although it remains the case that men's employment rates fell further than women's over the past few years. Future analysis should focus on demand-side factors, rather than assuming that most women either want to or are able to choose to stay at home.

The media hype about women opting out of employment is probably a result of the reality that for highly educated women, dropping out of the labor force is usually associated with having a child at home. What is interesting here is that just about the *only* reason that better-educated older women drop out of the labor force in the 2000s is to care for small children while at the same time, most highly educated women stay in the labor force when they have children.

References

Belkin, Lisa. 2003. "The Opt-Out Revolution," *New York Times Magazine*, October 26.

Boushey, Heather. 2005. *Family Friendly Policies: Helping Mothers to Make Ends Meet.* Washington, DC: Center for Economic and Policy Research.

Boushey, Heather, David Rosnick, and Dean Baker. 2005. *Gender Bias in the Current Economic Recovery? Declining Employment Rates for Women in the 21st Century*. Washington, DC: Center for Economic and Policy Research. Briefing Paper.

Story, Louise. 2005. "Many Women at Elite Colleges Set Career Path to Motherhood," *New York Times*, September 20.

Wallis, Claudia. 2004. "The Case for Staying Home," *Time*, May 10, p. 44.

Elite male faculty in the life sciences employ fewer women

Jason M. Sheltzer[a,1] and Joan C. Smith[b]

[a]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139; and [b]Twitter, Inc., Cambridge, MA 02139

Edited* by Shirley Tilghman, Princeton University, Princeton, NJ, and approved June 5, 2014 (received for review March 25, 2014)

Women make up over one-half of all doctoral recipients in biology-related fields but are vastly underrepresented at the faculty level in the life sciences. To explore the current causes of women's underrepresentation in biology, we collected publicly accessible data from university directories and faculty websites about the composition of biology laboratories at leading academic institutions in the United States. We found that male faculty members tended to employ fewer female graduate students and postdoctoral researchers (postdocs) than female faculty members did. Furthermore, elite male faculty—those whose research was funded by the Howard Hughes Medical Institute, who had been elected to the National Academy of Sciences, or who had won a major career award—trained significantly fewer women than other male faculty members. In contrast, elite female faculty did not exhibit a gender bias in employment patterns. New assistant professors at the institutions that we surveyed were largely comprised of postdoctoral researchers from these prominent laboratories, and correspondingly, the laboratories that produced assistant professors had an overabundance of male postdocs. Thus, one cause of the leaky pipeline in biomedical research may be the exclusion of women, or their self-selected absence, from certain high-achieving laboratories.

women in STEM | gender diversity

Between 1969 and 2009, the percentage of doctorates awarded to women in the life sciences increased from 15% to 52% (1, 2). Despite the vast gains at the doctoral level, women still lag behind in faculty appointments. Currently, only 36% of assistant professors and 18% of full professors in biology-related fields are women (3). The attrition of women from academic careers—known as the leaky pipeline problem (4)—undermines the meritocratic ideals of science and represents a significant underuse of the skills that are present in the pool of doctoral trainees.

A variety of factors has been suggested to influence the leaky pipeline in science, technology, engineering, and math (STEM) fields. Early career aspirations and choice of undergraduate major are significant departure points for women in certain disciplines (5, 6). For instance, women are awarded only 19% of bachelor's degrees in physics and 18% of bachelor's degrees in engineering, and correspondingly fewer women go on to graduate school in those subjects (1). In contrast, women are awarded >50% of both bachelor's and doctoral degrees in biology, suggesting that major leaks in the pipeline occur at later points in professional development. Gender differences in individuals' personal aspirations may explain some attrition from the academy (7). For instance, in surveys of graduate students and postdoctoral researchers (postdocs), women tend to rank work–life balance and parenthood-related issues as more important than men do, and the perceived difficulty of raising a family while working as a tenure-track faculty member causes more women than men to leave the academic pipeline (8–12). Such preferences are likely constrained by societal factors: male postdocs are more than two times as likely as female postdocs to expect their spouse to make career sacrifices for their benefit (8). Additionally, female scientists with children are significantly less likely to be hired for tenure-track jobs than those without children, whereas male scientists with children are more likely to be hired for tenure-track jobs than male scientists without children (13). Thus, a complex mixture of both free and constrained personal choices may contribute to the leaky pipeline in STEM fields.

In addition to the impact of gendered preference differences, the scarcity of female faculty may be, in part, because of persistent discrimination against women in science. Unlike systems of *de jure* discrimination, which were common until the middle of the 20th century and often explicitly excluded women from certain career paths, discrimination in the present day more often results from de facto differences in the treatment of men and women. Such behavior is linked to the problem of cumulative (dis)advantages: small differences in access to scientific goods (i.e., resources, mentoring, public visibility, etc.) may spiral over time, leading to significant divergence in achievement over the course of a career (14). These biases have been documented in both correlational and experimental studies of academic science. For instance, Moss-Racusin et al. (15) sent science faculty identical resumes for a laboratory manager position in which only the name and gender of the applicant were changed. The applicant with the male name was judged to be more competent and hirable and offered a larger starting salary than the female applicant.

How these gender biases affect the advancement of women in science is poorly understood. Moreover, in a field like biology—where women are well-represented at the doctoral and postdoctoral levels—it may be easy to assume that issues of gender are unimportant at early career stages. However, not all doctoral and postdoctoral positions are equivalent: vast interlaboratory differences exist in terms of reputation, mentoring, access to funding and equipment, networking possibilities, and more. Scientists who receive their training in particular laboratories may be at a disadvantage when applying for grants or faculty positions if their

Significance

Despite decades of progress, men still greatly outnumber women among biology faculty in the United States. Here, we show that high-achieving faculty members who are male train 10–40% fewer women in their laboratories relative to the number of women trained by other investigators. These skewed employment patterns may result from self-selection among female scientists or they may result from conscious or unconscious bias on the part of some faculty members. The dearth of women who are trained in these laboratories likely limits the number of female candidates who are most competitive for faculty job searches.

Author contributions: J.M.S. designed research; J.M.S. and J.C.S. performed research; J.C.S. contributed new reagents/analytic tools; J.M.S. and J.C.S. analyzed data; and J.M.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: sheltzer@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1403334111/-/DCSupplemental.

SOCIAL SCIENCES

principal investigator (PI) is less well-known or if their laboratory tends to produce fewer high-impact publications. We hypothesized that the steep decline in the representation of women at the postdoc-to-PI transition could be in part explained if the most prestigious PIs tended to predominantly train young male scientists. In this study we therefore sought to explore the link between gender, laboratory choice, and future academic employment. We found that women are significantly underrepresented at the graduate and postdoctoral levels in the laboratories of high-achieving male scientists, whereas elite female faculty show no such gender bias in their laboratories. We are unable to ascertain to what degree these differences result from self-selection among female trainees or gender biases among male faculty members. Nonetheless, this skew in laboratory employment represents a novel—and possibly corrigible—aspect of the leaky pipeline in life science research.

Results

A Survey of Employment by Gender at Top-Ranked Programs in the Life Sciences. To examine the gender distribution of biomedical scientists in academia, we collected information on the graduate students, postdocs, and faculty employed in 39 departments at 24 of the highest-ranked research institutions in the United States (*SI Materials and Methods* and Table S1). We focused on departments that study molecular biology, cell biology, biochemistry, and/or genetics. For each faculty member, we determined their gender, academic rank, and if available, the year in which their doctorate was received. We then used their laboratory website or a departmental directory to find the names of graduate students and postdocs currently working in their laboratories. We attempted to assign a gender to each graduate student and postdoc, using the internet and social network searches when a name was ambiguous. Lastly, we used three different criteria to define faculty whose laboratories we hypothesized would be the most prestigious: those who were funded by the Howard Hughes Medical Institute (HHMI), were members of the National Academy of Sciences (NAS), or had won at least one of seven different major research awards (e.g., the Nobel Prize or the National Medal of Science) (*SI Materials and Methods* and Table S2).

In total, we obtained information on 2,062 faculty members in the life sciences (Table 1). Within this sample, 21% of full professors and 29% of assistant professors were women. The gender distribution of the faculty members that we classified as elite was approximately proportional to the gender distribution of full professors, as 25% of HHMI investigators, 18% of NAS members, and 16% of major award winners were female. Among the trainees that we counted, 49% of 4,143 graduate students and 39% of 4,904 postdocs were women. We note that the percentages of female postdocs and assistant professors in our sample are below national averages (39% vs. 43% and 29% vs. 36%, respectively), suggesting that the leaky pipeline might be comparatively worse at top-ranked research institutions (3, 16). Alternately, the subfields from which we harvested might be slightly more male-biased than other fields, like ecology, that also fall under the spectrum of the life sciences (16).

Table 1. A survey of biology laboratories in the United States

Category	Male	Female	Total
Faculty	1,557	505	2,062
Professor	1,023	276	1,299
Associate professor	269	121	390
Assistant professor	265	108	373
HHMI investigators	113	38	151
NAS members	210	47	257
Major award winners	53	10	63
Faculty with one or more trainees listed	982	358	1,340
Postdoctoral researchers	3,013	1,891	4,904
Graduate students	2,120	2,023	4,143

Elite Male Faculty Employ Fewer Female Graduate Students and Postdocs. We next examined the gender distribution of trainees on a per-laboratory basis. On average, male PIs ran laboratories that had 36% female postdocs and 47% female graduate students (Fig. 1*A*). These values were significantly lower than values that we observed in laboratories headed by women, who employed on average 46% female postdocs and 53% female graduate students (Fig. 1*B*) ($P < 0.0001$ for both comparisons, Wilcoxon rank sum test). Thus, male professors run laboratories that have about 22% fewer female postdocs and 11% fewer female graduate students than their female colleagues do (Fig. 2).

We then tested our data to determine if the most prestigious PIs trained fewer women. Surprisingly, for each comparison, male PIs who were funded by HHMI, were elected to the NAS, or had won a major research award employed significantly fewer female postdocs than the corresponding pool of other male PIs (Fig. 1*A*). For instance, male HHMI investigators ran laboratories that had, on average, 31% female postdocs, whereas men who were not HHMI investigators employed, on average, 38% female postdocs ($P < 0.0001$, Wilcoxon rank sum test). This difference translates to a 19% deficit in the employment of female postdocs relative to their representation across all laboratories (Fig. 2*A*). In contrast, female professors who had achieved the same career milestones showed no evidence of a gender bias. Women who were HHMI investigators ran laboratories that employed 48% female postdocs compared with 46% female postdocs among those who were not funded by the HHMI (Fig. 1*B*). Similar results were obtained when we examined women who were members of the NAS or had won a major research award.

At the graduate student level, we observed an analogous, although less substantial, skew in employment by gender. Male NAS members and major award winners ran laboratories with about 41–42% female graduate students compared with 47–48% among other male professors (Fig. 1*A*). This disparity represents 14% and 17% deficits in the employment of female graduate students by NAS and award-winning laboratories, respectively, relative to their representation across all laboratories (Fig. 2*B*). However, there was no difference at the graduate student level between HHMI- and non-HHMI–funded male PIs (Fig. 1*A*). Among female faculty, major award winners actually trained slightly more female graduate students than non-award winners, whereas HHMI funding and NAS membership did not affect the number of female graduate students employed by female professors (Fig. 1*B*). Thus, elite male PIs, but not elite female PIs, tend to employ fewer female trainees than other faculty members who have not achieved certain career milestones.

Our dataset also included 24 Nobel Laureates in Medicine/Physiology or Chemistry. Male PIs who had won a Nobel Prize ($n = 22$) ran laboratories that had, on average, 24% female postdocs and 36% female graduate students, which represents a 39% and 27% deficit, respectively, relative to the pool of trainees (Fig. S1). The paucity of female Nobel Laureates prevented a meaningful comparison using this criterion, although we note that both female Nobel Laureates in our sample ran laboratories in which female trainees outnumbered male trainees at the time of our survey.

These results led us to consider the distribution of trainees across all laboratory types. We found that female trainees were much less likely to work for an elite PI, particularly at the postdoctoral level (Fig. S2). Combining faculty of both genders, men were about 17% more likely to do their graduate training with a member of the NAS, 25% more likely to do their postdoctoral training with a member of the NAS, and 90% more likely to do

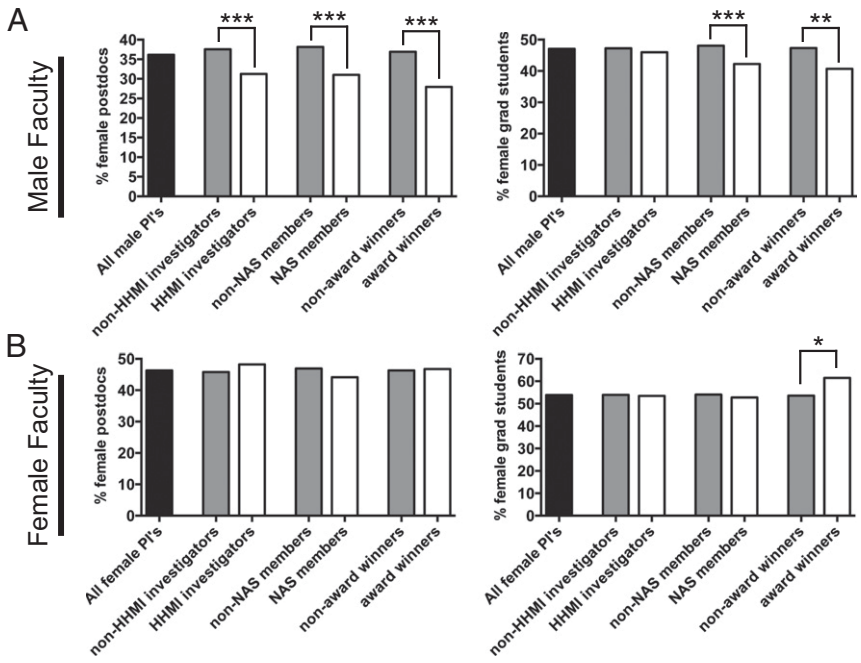


Fig. 1. The gender composition of elite biology laboratories in the United States. The weighted average percentages of female trainees in laboratories with (*A*) male PIs and (*B*) female PIs who have achieved certain career milestones are displayed. Major career awards that were counted for this survey are listed in Table S2. $*P < 0.05$; $**P < 0.005$; $***P < 0.0005$ (Wilcoxon rank sum test).

their postdoctoral training with a Nobel Laureate. Thus, the gender skew in employment results in fewer women being trained in the laboratories of elite investigators.

Other Factors That Affect the Gender Skew in Biology Laboratories. We next sought to identify other characteristics of PIs that correlate with altered gender distributions. Nearly every faculty member who had achieved one of the career milestones that we counted held the rank of professor; thus, our results could be explained if older faculty in general trained few women. In fact, men who were full professors tended to employ fewer female postdocs but more female graduate students than men who were assistant professors (Fig. S3). Women who were full professors also trained fewer female postdocs than assistant professors did, but there was no difference in graduate student employment between women with different academic ranks. Nonetheless, when we restricted our analysis to only faculty holding the rank of full professor, we still observed a significant deficit in women trained specifically by elite male PIs (Fig. S3).

Among different STEM disciplines, the representation of women generally decreases in more math-intensive fields (3, 17). We focused our analysis on departments of cell and molecular biology (Table S1), but our dataset does contain biophysicists, computational biologists, and other investigators who take a more quantitative approach to biological questions. HHMI investigators whose listed discipline was biophysics, computational biology, or systems biology and NAS members whose primary section was Biophysics and Computational Biology were found to employ particularly few female trainees (Fig. S4). However, even with these faculty members excluded, the remaining male HHMI investigators and NAS members trained fewer women than other male PIs did (Fig. S4).

Lastly, we built various linear models from our dataset using either the weighted percentage of female postdocs or the weighted percentage of female graduate students in each laboratory as the dependent variable. For simplicity, we collapsed the three career achievements that we scored into a single categorical variable (elite status). As expected, among male faculty, elite status was negatively correlated with the percentage of female postdocs in a laboratory ($P < 0.0001$) (Table S3). This relationship remained true even when several other explanatory variables were added, including faculty rank, years since a faculty member had received his or her PhD, and total number of trainees in a laboratory (Table S4). As a single independent variable, years since PhD was moderately negatively correlated with the percentage of

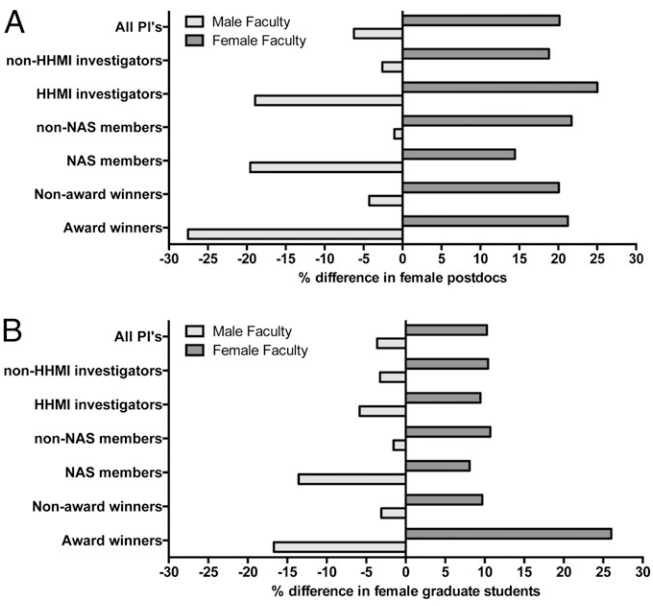


Fig. 2. Elite male PIs employ fewer women. The percent differences in (*A*) female postdocs and (*B*) female graduate students employed by PIs who have achieved certain career milestones are displayed. The axis at $x = 0$ represents employing female trainees at a rate proportional to their representation among all laboratories in this survey.

female postdocs in laboratories with male faculty members ($P < 0.045$), but this effect disappeared when other variables were included in the model (Tables S3 and S4). This observation suggests that a faculty member's age is not a significant determinant of the gender makeup of their laboratory, and both young and old elite professors employ few women. Laboratory size was also negatively correlated with the representation of female postdocs both as a single variable and in multivariable models. Regression against the percentage of female graduate students in each laboratory revealed similar, although less robust, results. In multivariable models, elite status was associated with a significantly lower percentage of female graduate students trained by male faculty (Table S4). However, years since PhD correlated with an increasing representation of female graduate students, whereas laboratory size was not significantly correlated in either direction. Finally, we constructed equivalent linear models for female PIs, but we failed to find a single variable that was significantly associated with differential representation of female trainees in these laboratories.

Laboratories That Produce Assistant Professors Employ More Male Postdocs. In the current funding environment, there is intense competition among postdocs for scarce tenure-track positions as assistant professors (18, 19). We sought to determine how postdoctoral laboratory choice and gender influenced this process. Using curriculum vitae, websites, and publication records, we determined the prior employment of 311 of 373 assistant professors from the 39 departments that we surveyed. Of these, 276 assistant professors were postdocs before their faculty appointments, and 144 of them completed their postdoc in 1 of 118 laboratories that we had surveyed. (Assistant professors who did not complete postdocs primarily held clinical positions or were independent fellows before their faculty appointments.) Accordingly, we examined the characteristics of these feeder laboratories that have successfully trained postdocs who won recent faculty job searches at top universities.

The PIs of feeder laboratories were significantly more likely to be HHMI investigators, NAS members, or have won a major research award relative to the pool of all PIs (Fig. 3*A*). For instance, 13% of the professors in our dataset were members of the NAS, but 58% of feeder laboratory professors were NAS members ($P < 0.0001$, Fisher exact test). Our above analysis suggested that these feeder laboratories may therefore have skewed gender ratios. Indeed, these laboratories had 14% fewer female postdocs than nonfeeder laboratories did (Fig. 3*B*) ($P < 0.0001$, Wilcoxon rank sum test). Both male and female faculty, considered separately, who had trained new professors employed fewer female postdocs than those who had not, although for female faculty, the percentage was still higher than the representation of female postdocs across all laboratories (Fig. 3 *B–D*). Because 71% of the assistant professors in our sample were male, these results could represent a form of selection bias. However, when we examined only the feeder laboratories in which female assistant professors had trained, we found that these laboratories still employed disproportionately few female postdocs (Fig. 3*B*). We conclude that PIs who successfully train new assistant professors employ an overabundance of male postdocs.

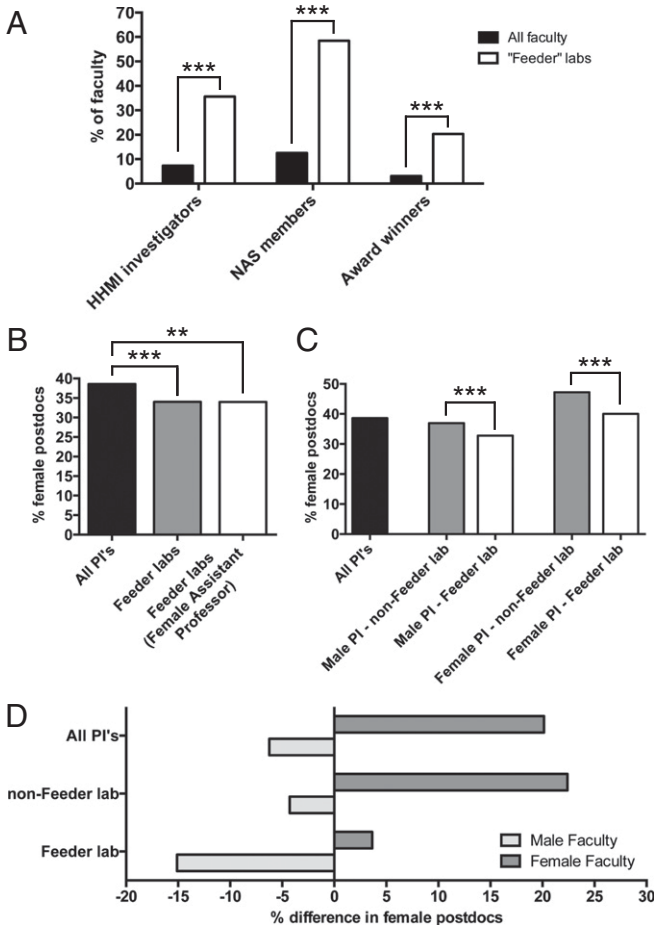


Fig. 3. Feeder laboratories train fewer female postdocs. (*A*) The percentages of all faculty members (black bars) and faculty members who have recently trained new assistant professors (white bars) who have achieved certain career milestones are displayed. ***$P < 0.0005$ (Fisher exact test). (*B*) The percentages of female postdocs employed in different laboratories are displayed. The black bar represents all PIs, the gray bar represents PIs who have recently trained at least one new assistant professor of either gender, and the white bar represents PIs who have recently trained at least one new female assistant professor. **$P < .005$; ***$P < 0.0005$ (Fisher exact test). (*C*) The percentage of female postdocs employed in different laboratory types subdivided by the gender of the PI is displayed. ***$P < 0.0005$ (Fisher exact test). (*D*) The percentage differences in female postdocs employed by different PIs are displayed. The axis at $x = 0$ represents employing female postdocs at a rate proportional to their representation among all laboratories in this survey.

Discussion

Our results show that male faculty in general, and elite male faculty in particular, train fewer female graduate students and postdocs relative to their representation in the pool of trainees at top universities. These findings are robust to considerations of faculty rank, age, and laboratory size and cannot be explained by the exclusion of women from a small number of math-intensive laboratories. Because the majority of assistant professors in our dataset conducted their postdoctoral research under the supervision of one of these high-achieving PIs, the limited number of women trained in these laboratories reduces the number of female candidates who would be most competitive for faculty job searches.

Notably, our current data do not show conscious bias on the part of male PIs who employ few female graduate students and postdocs. It may be the case that women apply less frequently to laboratories with elite male PIs. Unfortunately, data on this question are difficult to collect, because applying to do research in a laboratory is an unregulated and largely informal process. Interested graduate students and postdoctoral candidates typically e-mail PIs with whom they seek to work, and faculty members can easily ignore or delete requests when they so choose. Milkman et al. (20) have shown that faculty members across a host of disciplines, including the life sciences, respond to e-mails from prospective graduate students written under male names significantly more frequently than they respond to e-mails written under female names. Nonetheless, self-selection among female

graduate students and postdocs may still contribute to the gender skew that we have documented. By graduate school, fewer women than men perceive themselves to be on an academic career track (10, 21). Some women, in particular, may not apply to the most prestigious laboratories if they do not believe it to be important for their professional development. Additionally, in certain circumstances, women underrate their own skill sets (22–24), which could lead some to self-select away from elite laboratories.

A lack of applications from women could also reflect specific issues with a laboratory or PI. Female PhDs frequently cite marriage and childbirth as reasons to opt out of scientific careers (12); faculty members who are reputed to be hostile toward maternity considerations could be implicitly discouraging women from applying to their laboratories. More insidiously, 16% of women employed in the academy report that they have experienced work-related sexual harassment (25), and the tolerance of sexual harassment in a laboratory could further decrease the number of female applicants. Regardless of the cause, we believe that PIs who receive few applications from female trainees ought to increase their efforts to proactively recruit talented women and should ensure that their laboratories are safe spaces for female scientists. A more formalized process of applying to work as a graduate student or postdoc could also serve as a check against PIs who routinely fail to hire women.

Alternately, differences in quality between the male and female applicant pools could contribute to the gender gap in elite laboratories with male PIs. We note, however, that female investigators at the top of their respective fields run laboratories with just as many women as other female PIs (Fig. 1*B*). Additionally, women win competitive fellowships for graduate and postdoctoral training at frequencies that are proportional to their representation among all trainees: about 55% of National Science Foundation graduate fellowships, 45% of Helen Hay Whitney postdoctoral fellowships, and 41% of Jane Coffin Childs postdoctoral fellowships are awarded to women (26–28). These observations argue against a sizeable gap in applicant quality between male and female trainees in the life sciences.

Thus, in addition to the aforementioned factors, we suggest that gender bias may contribute to the decreased employment of women in laboratories with elite male PIs. Several recent studies have shown that gender bias remains an endemic problem in academic science. For instance, in several European countries, faculty promotion decisions are made by randomly chosen review committees. The promotion chances of female candidates are significantly decreased if they are assigned to an all-male review committee, whereas their promotion chances are equivalent or nearly equivalent to men's chances if they are assigned to a mixed-gender committee (29, 30). Similar results have been reported in the private sector: men tend to underrate women's job performances (31), whereas having more women in supervisory roles is associated with the increased hiring and promotion of other women (32). In an academic context, graduate student and postdoc employment decisions are often made unilaterally by the single PI who runs a laboratory. It may be the case that, in the most competitive laboratories, male PIs knowingly or unknowingly underestimate the qualifications of female applicants and instead hire more men in their place. Because mixed-gender hiring committees more accurately assess the qualifications of female applicants in a variety of settings (29–31), male PIs could potentially benefit by soliciting feedback on applicants to their laboratories from female postdocs or affiliated faculty members.

Irrespective of the cause of the gender disparities in elite laboratories, its consequences significantly shape the academic ecosystem. Our data show that these laboratories function as gateways to the professoriate: new generations of faculty members are predominantly drawn from postdocs trained by high-achieving PIs. However, these feeder laboratories employ a disproportionate number of men (Fig. 3). According to the theory of cumulative disadvantage, persistent inequalities in achievement can result from small differences in treatment over a prolonged goal-oriented process (14). In controlled studies, women in academia receive less favorable evaluations, receive lower salary offers, and are ignored by faculty more frequently than men (15, 20). Access to training in certain laboratories may be another level at which women are disadvantaged. The absence or exclusion of female trainees from elite laboratories deprives them of the resources, visibility, networking opportunities, etc. that could facilitate their professional development. These differences may contribute to the leaky pipeline by shunting women toward laboratories that provide fewer opportunities for advancement in academic science.

Conclusions

The continued underrepresentation of women in the academy slows the progress of discovery by artificially excluding individuals with the ability to make significant contributions to the scientific enterprise. It is our hope that this work, along with the growing body of related evidence showing gender bias in the academy (15, 20, 29, 30), will elicit an increased awareness of the ways in which gender continues to play a role in shaping the career trajectories of young scientists. Recognition of gender disparities as a persistent problem can aid in the fair evaluation of women in hiring decisions and trigger active steps by individual PIs to recruit more talented women to their laboratories. Such steps can ensure that, in the future, an individual's gender will not hinder their ability to engage in scientific research.

Materials and Methods

The name, gender, and rank of PIs from 39 biology departments at 24 academic institutions were collected (Table S1). Graduate students and postdocs who worked for each faculty member were identified from public departmental listings or their laboratory's website. Current HHMI investigators were downloaded from www.hhmi.org and matched with faculty in our database. Members of the NAS were downloaded from www.nasonline.org and matched with faculty in our database. Winners of major research awards (listed in Table S2) through 2013 were downloaded from their respective websites and matched with faculty in our database. Additional details on our survey methodology are presented in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Angelika Amon, Terry Orr-Weaver, Frank Solomon, Kendra Albert, and members of the Amon laboratory for helpful comments on the manuscript.

1. National Science Board (2012) *Science and Engineering Indicators 2012*. Available at http://www.nsf.gov/statistics/seind12/. Accessed November 1, 2013.
2. National Science Foundation, Division of Science Resources Statistics (2008) *Thirty-Three Years of Women in S&E Faculty Positions*. Available at www.nsf.gov/statistics/infbrief/nsf08308/. Accessed January 20, 2014.
3. Nelson DJ (2007) *A National Analysis of Minorities in Science and Engineering Faculties at Research Universities*. Available at http://faculty-staff.ou.edu/N/Donna.J.Nelson-1/diversity/Faculty_Tables_FY07/FinalReport07.html. Accessed January 20, 2014.
4. Pell AN (1996) Fixing the leaky pipeline: Women scientists in academia. *J Anim Sci* 74(11):2843–2848.
5. Sadler PM, Sonnert G, Hazari Z, Tai R (2012) Stability and volatility of STEM career interest in high school: A gender study. *Sci Educ* 96(3):411–427.
6. Morgan SL, Gelbgiser D, Weeden KA (2013) Feeding the pipeline: Gender, occupational plans, and college major selection. *Soc Sci Res* 42(4):989–1005.
7. Ceci SJ, Williams WM (2011) Understanding current causes of women's underrepresentation in science. *Proc Natl Acad Sci USA* 108(8):3157–3162.
8. Martinez ED, et al. (2007) Falling off the academic bandwagon. Women are more likely to quit at the postdoc to principal investigator transition. *EMBO Rep* 8(11):977–981.
9. Ferriman K, Lubinski D, Benbow CP (2009) Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging adulthood and parenthood. *J Pers Soc Psychol* 97(3):517–532.
10. van Anders SM (2004) Why the academic pipeline leaks: Fewer men than women perceive barriers to becoming professors. *Sex Roles* 51(9-10):511–521.

11. Massachusetts Institute of Technology (2011) *Postdoctoral Life at MIT. Findings from the 2010 Postdoctoral Scholar Survey*. Available at http://web.mit.edu/ir/surveys/pdf/Postdoctoral_Life_at_MIT_Report_June_2011.pdf. Accessed January 19, 2014.
12. Goulden M, Mason MA, Frasch K (2011) Keeping women in the science pipeline. *Ann Am Acad Pol Soc Sci* 638(1):141–162.
13. Ginther DK, Kahn S (2006) *Does Science Promote Women? Evidence from Academia 1973–2001 (National Bureau of Economic Research)*. Available at http://www.nber.org/papers/w12691. Accessed February 1, 2014.
14. DiPrete TA, Eirich GM (2006) Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annu Rev Sociol* 32:271–297.
15. Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J (2012) Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci USA* 109(41):16474–16479.
16. National Science Foundation, National Center for Science and Engineering Statistics (2013) *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2013*. Available at http://www.nsf.gov/statistics/wmpd/. Accessed January 15, 2014.
17. Ceci SJ, Williams WM (2010) Sex differences in math-intensive fields. *Curr Dir Psychol Sci* 19(5):275–279.
18. Carpenter S (2010) *Tenure-Track Jobs Remain Scarce. Science*. Available at http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2010_01_15/caredit.a1000006. Accessed February 7, 2014.
19. Price M (2013) *Excessive Supply, Uncertain Demand. Science*. Available at http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2013_09_16/caredit.a1300199. Accessed February 7, 2014.
20. Milkman KL, Akinola M, Chugh D (2013) *Discrimination in the Academy: A Field Experiment (Social Science Research Network, Rochester, NY)*. Available at http://papers.ssrn.com/abstract=2063742. Accessed January 6, 2014.
21. Fox MF, Stephan PE (2001) Careers of young scientists: Preferences, prospects and realities by gender and field. *Soc Stud Sci* 31(1):109–122.
22. Steinmayr R, Spinath B (2009) What explains boys' stronger confidence in their intelligence? *Sex Roles* 61:736–749.
23. Correll SJ (2001) Gender and the career choice process: The role of biased self-assessments. *AJS* 106(6):1691–1730.
24. Pallier G (2003) Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles* 48(5-6):265–276.
25. Ilies R, Hauserman N, Schwochau S, Stibal J (2003) Reported incidence rates of work-related sexual harassment in the United States: Using meta-analysis to explain reported rate disparities. *Person Psychol* 56(3):607–631.
26. Helen Hay Whitney Foundation (2013) *Current Fellowship Recipients*. Available at http://www.hhwf.org/HTMLSrc/Directory.html. Accessed February 11, 2014.
27. National Science Foundation (2013) *2013 Graduate Research Fellowships Reflect a Diversity of Fields, Institutions and Students*. Available at http://www.nsf.gov/news/news_summ.jsp?cntn_id=127538. Accessed February 21, 2014.
28. Jane Coffin Childs Memorial Fund for Medical Research (2014) *The Jane Coffin Childs Fund Fellows 2011–2014*. Available at http://www.jccfund.org/about-fund/fellows. Accessed February 21, 2014.
29. Zinovyeva N, Bagues M (2010) *Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment (Social Science Research Network, Rochester, NY)*. Available at http://papers.ssrn.com/abstract=1618256. Accessed February 11, 2014.
30. De Paola M, Scoppa V (2011) *Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy (Università della Calabria, Dipartimento di Economia, Statistica e Finanza (Ex Dipartimento di Economia e Statistica))*. Available at http://econpapers.repec.org/paper/clbwpaper/201106.htm. Accessed February 11, 2014.
31. Bowen C-C, Swim JK, Jacobs RR (2000) Evaluating gender biases on actual job performance of real people: A meta-analysis. *J Appl Soc Psychol* 30(10):2194–2215.
32. Kurtulus FA, Tomaskovic-Devey D (2012) Do female top managers help women to advance? A panel study using EEO-1 records. *Ann Am Acad Pol Soc Sci* 639(1):173–197.