

National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track

Wendy M. Williams¹ and Stephen J. Ceci

Department of Human Development, Cornell University, Ithaca, NY 14853

Edited* by Richard E. Nisbett, University of Michigan, Ann Arbor, MI, and approved March 5, 2015 (received for review September 30, 2014)

National randomized experiments and validation studies were conducted on 873 tenure-track faculty (439 male, 434 female) from biology, engineering, economics, and psychology at 371 universities/colleges from 50 US states and the District of Columbia. In the main experiment, 363 faculty members evaluated narrative summaries describing hypothetical female and male applicants for tenure-track assistant professorships who shared the same lifestyle (e.g., single without children, married with children). Applicants' profiles were systematically varied to disguise identically rated scholarship; profiles were counterbalanced by gender across faculty to enable between-faculty comparisons of hiring preferences for identically qualified women versus men. Results revealed a 2:1 preference for women by faculty of both genders across both math-intensive and non-math-intensive fields, with the single exception of male economists, who showed no gender preference. Results were replicated using weighted analyses to control for national sample characteristics. In follow-up experiments, 144 faculty evaluated competing applicants with differing lifestyles (e.g., divorced mother vs. married father), and 204 faculty compared same-gender candidates with children, but differing in whether they took 1-y-parental leaves in graduate school. Women preferred divorced mothers to married fathers; men preferred mothers who took leaves to mothers who did not. In two validation studies, 35 engineering faculty provided rankings using full curricula vitae instead of narratives, and 127 faculty rated one applicant rather than choosing from a mixed-gender group; the same preference for women was shown by faculty of both genders. These results suggest it is a propitious time for women launching careers in academic science. Messages to the contrary may discourage women from applying for STEM (science, technology, engineering, mathematics) tenure-track assistant professorships.

gender bias | hiring bias | underrepresentation of women |
faculty hiring | women in science

Women considering careers in academic science confront stark portrayals of the treacherous journey to becoming professors. Well-publicized research depicts a thicket of obstacles standing between female graduate students and tenure-track positions, including inadequate mentoring and networking (1); a chilly social climate (2); downgrading of work products such as manuscripts (3), grant proposals (4), and lectures (5); and gender bias in interviewing and hiring (6–9). Numerous blue ribbon panels and national reports have concluded that implicit, and sometimes explicit, attitudes pervade the hiring process and negatively influence evaluations of female candidates and their scholarship, contributing to women's underrepresentation within the academy (e.g., refs. 10–13).

Women's underrepresentation in academic science is hardly trivial. In life and social sciences, women now earn the majority of doctorates, but they make up a minority of assistant professors. In 1993–1995, 28.4% of assistant professors were women, but 41.6% of Ph.D.s awarded in the same cohort went to women. That is, almost one-third of the women did not advance from receiving their Ph.D. to an assistant professorship (see ref. 14, figure 5). More recently, in 2008–2010, this gap widened to 22 percentage points (53.2% of doctorates to women; 31.6% of assistant professorships to women), and this gap persisted after

controlling for demographics, degree characteristics, and field (15). [This winnowing of women in the STEM (science, technology, engineering, mathematics) tenure-track pipeline is a result of women Ph.D.s being far less likely than men to apply for tenure-track jobs, rather than to women applying but being rejected at higher rates than men (14).] Against this bleak backdrop, it is perhaps no surprise that talented young women opt out of the STEM tenure track either by not applying for assistant professorships at the same rate as men or, in some fields, by not majoring in them in college in the first place (14).

The point at which scientists choose to apply for tenure-track assistant professorships is a key juncture in understanding the problem of women's underrepresentation. Once hired, women prosper in the STEM professoriate (14, 16–18): They are remunerated, persist, and are promoted at rates roughly comparable to men's (14) after controlling for observable characteristics, including academic productivity. However, to be hired and eventually tenured, women must first apply. Unfortunately, despite their success once hired, women apply for tenure-track positions in far smaller percentages than their male graduate student counterparts (14, 16, 18). Why might this be?

One reason may be omnipresent discouraging messages about sexism in hiring, but does current evidence support such messages? Despite this question's centrality to any informed discussion about women's underrepresentation in academic science, only one experimental study (7) contrasted faculty ratings of the relative "hirability" of hypothetical identically qualified women and men. Results showed that both female and male psychology faculty members downgraded a hypothetical woman's academic record compared with an identical man's. However, this study

Significance

The underrepresentation of women in academic science is typically attributed, both in scientific literature and in the media, to sexist hiring. Here we report five hiring experiments in which faculty evaluated hypothetical female and male applicants, using systematically varied profiles disguising identical scholarship, for assistant professorships in biology, engineering, economics, and psychology. Contrary to prevailing assumptions, men and women faculty members from all four fields preferred female applicants 2:1 over identically qualified males with matching lifestyles (single, married, divorced), with the exception of male economists, who showed no gender preference. Comparing different lifestyles revealed that women preferred divorced mothers to married fathers and that men preferred mothers who took parental leaves to mothers who did not. Our findings, supported by real-world academic hiring data, suggest advantages for women launching academic science careers.

Author contributions: W.M.W. and S.J.C. designed research, performed research, analyzed data, and 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.

¹To whom correspondence should be addressed. Email: wendywilliams@cornell.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1418878112/-DCSupplemental.

was published 16 y ago and involved only one field, psychology, a discipline that is more than 50% female (14).

There are two critical omissions from the current data landscape. First, no experimental study of tenure-track faculty hiring in math-intensive fields has ever evaluated whether bias can explain women's underrepresentation in those fields today. This is important because it is in math-intensive fields that women are most underrepresented (see ref. 14 for historical and current data). Second, no current experimental study demonstrating sexist hiring is based on actual faculty members who were asked to rate identically qualified tenure-track assistant professor applicants. Instead, past studies have used ratings of students' hirability for a range of posts that do not include tenure-track jobs, such as managing laboratories or performing math assignments for a company. However, hiring tenure-track faculty differs from hiring lower-level staff: it entails selecting among highly accomplished candidates, all of whom have completed Ph.D.s and amassed publications and strong letters of support. Hiring bias may occur when applicants' records are ambiguous, as was true in studies of hiring bias for lower-level staff posts, but such bias may not occur when records are clearly strong, as is the case with tenure-track hiring. Thus, we focused on male and female assistant professor candidates who were identically and unambiguously strong and used tenured/tenure-track STEM faculty members from all 50 US states as raters to determine the role of gender bias in tenure-track faculty hiring.

This program of research consisted of five experiments involving 873 tenure-track faculty members from 371 colleges and universities, spanning 50 states and the District of Columbia. We investigated faculty hiring preferences for hypothetical applicants in two math-intensive fields in which women are substantially underrepresented (engineering and economics) and two non-math-intensive fields in which women are well represented (biology and psychology; see ref. 14, figure 4A, for current and historical data on women in these fields). We used (with embellishments) a method for revealing gender bias in hiring that has been used frequently in past studies (6, 7, 19): We compared the likelihood of identically qualified women and men (named Dr. X, Dr. Y., and Dr. Z, and differing across faculty raters solely in the gender pronoun used to refer to them) being ranked first by individual faculty members for a tenure-track assistant professorship (*SI Appendix*).

To illuminate contextual factors in faculty decision-making, we also studied the effects of candidate lifestyle on hiring preference for otherwise-identical male and female candidates. For example, in experiment 1, we included two applicants, both of whom were unmarried and childless or, in another condition, both of whom were married with preschool-age children and stay-at-home spouses, to see how a candidate's perceived hirability was influenced by various lifestyle situations. In experiment 2, which focused on nonmatching lifestyles, we included a married father of two with a stay-at-home spouse competing against a single mother of two with an absent ex-spouse (in addition to a third, "foil," candidate appearing in every contest, as described later). Because so much research points to fertility decisions as key determinants of women's decisions to opt out (20–23), in experiment 3 we explored the relative hirability of identical mothers (or in a separate condition, identical fathers) who either took or did not take a 1-y parental leave in graduate school. [Women's perceptions that an extended maternity leave will cause them to be viewed as less committed to their profession (22, 23) may influence some women to opt out entirely.] Experiment 4 used the same method as experiment 1, except faculty members rated candidates' full curricula vitae (CVs), rather than narrative summaries. In experiment 5, faculty members rated only a single applicant (female or male, with identical records) to ensure the findings of experiments 1–4 were not a result of socially desirable responding.

Our research required a design enabling a realistic comparison of a woman and man with identical credentials applying for the same assistant professorship without signaling our key hypotheses

to faculty. We could not simply send faculty members two identical candidate descriptions differing only in gender and ask which person the faculty member preferred to hire. Such a transparent approach would have revealed our central question and compromised the results. Thus, the design of experiments 1–4 was necessarily complex, relying on comparisons among faculty members to reveal how hiring preferences were influenced by candidate gender and lifestyle. Faculty members received information regarding three short-listed candidates for a tenure-track assistant professorship in their department. This information included the search committee chair's notes reflecting the committee's evaluation of each candidate's scholarly record (based on having read the vita and actual publications); the information also included excerpts from letters of reference and the faculty's average ratings of the candidate's job talk and interview. Narrative summaries of candidate credentials rather than CVs were used because of our sampling over four fields and three major Carnegie-institution levels while needing to simultaneously hold constant applicant credentials (i.e., no single vita would be realistic across both large, research-intensive universities and small colleges). In addition, the chair's comments about fit with the department were noted: "Dr. Z: Z struck the search committee as a real powerhouse. Based on her vita, letters of recommendation, and their own reading of her work, the committee rated Z's research record as 'extremely strong.' Z's recommenders all especially noted her high productivity, impressive analytical ability, independence, ambition, and competitive skills, with comments like, 'Z produces high-quality research and always stands up under pressure, often working on multiple projects at a time.' They described her tendency to 'tirelessly and single-mindedly work long hours on research, as though she is on a mission to build an impressive portfolio of work.' She also won a dissertation award in her final year of graduate school. Z's faculty job talk/interview score was 9.5/10. At dinner with the committee, she impressed everyone as being a confident and professional individual with a great deal to offer the department. During our private meeting, Z was enthusiastic about our department, and there did not appear to be any obstacles if we decided to offer her the job. She said her husband is a physician (with opportunities to practice in our area) and that they will need two slots at university daycare for their children. Z said our department has all of the resources needed for her research." (In the full-vita condition, experiment 4, faculty members evaluated actual full CVs of engineering applicants in place of summaries.) We asked faculty members to rank the candidates first, second, and third for the job.

Twenty counterbalanced sets of materials for experiments 1–3 were developed in which the same candidate was depicted as "she" in one version and "he" in the next. We also varied lifestyle for candidates in the same way, using identical candidate records of accomplishments but varying descriptors such as "single with no children," "married, needs two slots in university daycare," or "divorced, needs two slots in university daycare": disclosures portrayed in our materials as having been voiced spontaneously by candidates themselves, as sometimes happens during job interviews. One tactic to disguise the hypothesis and increase the realism of the rating task was the use of a pretested foil candidate in each set of three candidates. The foil candidate was created to be slightly weaker than the other two candidates but still excellent. The foil rounded out a slate of three candidates and allowed us (in certain experimental conditions) to present two male and one female candidate to faculty, which was important in creating a realistic short-list of job finalists, especially in disciplines such as engineering and economics, which have far fewer women than men applicants. Another disguise tactic was the incorporation of ecologically rich personal descriptions of candidates that included their demeanor at the search committee dinner. Two forms of these descriptions were developed: one with traditionally female adjectives ("imaginative, highly creative, likeable, kind, socially skilled") and one with traditionally male adjectives ("analytical, ambitious, independent, stands up under pressure, powerhouse") (24, 25). In every case, actual

candidate gender was counterbalanced with adjective gender across versions, so that half of the faculty members were sent a given candidate depicted as a woman described with male adjectives and half were sent the same woman depicted with female adjectives, and vice versa for male candidates.

In all 20 sets of materials in experiments 1–3, the contest was between Drs. X, Y, and Z (with Y being the foil candidate). Faculty members were sent emails containing one of the 20 versions and asked to rank the three candidates for an assistant professorship in their department. Versions were sent out randomly after stratification of the sample by the three major Carnegie classifications of institutions (doctoral-granting universities, master's colleges and universities, and baccalaureate colleges), department (biology, engineering, economics, psychology), and faculty gender. The 20 versions of materials allowed us to compare rankings between faculty raters of Dr. X as a woman with rankings of Dr. X as a man. Again, both had identical credentials and were described with identical adjectives; the sole difference was the gender pronoun used to describe Dr. X: “he” in one version and “she” in the other. The same manipulation was performed for Dr. Z. (Remember that Dr. Y was the foil, pretested to be slightly weaker than Dr. X and Dr. Z, but still excellent; the foil was ranked first 2.53% of the time.)

Our method enabled us to contrast the hirability of women and men candidates in a between-subjects design, with counterbalancing of candidate gender and adjective gender between faculty members (half female, half male), so that rankings of identically qualified candidates of different genders and lifestyles could be compared across 873 faculty from 371 institutions in 50 US states. Multiple validations (described here and in the *SI Appendix*) were incorporated into our design, including use of full vitas to ensure results matched those for vita summaries, use of a paid sample with 91% response rate to validate results from the national sample, and validation of adjectives to ensure past findings of gender connotations applied to faculty in our study. Our experimental design thus disentangled the relative effects on hiring preferences of applicant gender, applicant lifestyle, faculty gender, and faculty field.

Results and Discussion

In experiments 1–3, 20 sets of materials were sent to 2,090 faculty members across the United States, half female and half male; 711 voluntarily ranked applicants (34.02%). (Cornell University's institutional review board approved this study; faculty were free to ignore our emailed survey.) The response rates for every cell (university Carnegie type by department by gender, $3 \times 4 \times 2$) were evaluated in a logistic regression and shown to be unrelated to the findings. Our analyses examined which candidate was ranked first, under which conditions, by faculty of each gender and field. We analyzed the data using two independent approaches [traditional unweighted analysis (simple random sample, reported herein) and weighted analysis] to validate the generalizability of our findings. The unweighted analysis used logistic regression to predict faculty rankings of candidates. The weighted analysis assigned a sample weight to each faculty member on the basis of the numbers of women/men in her/his academic department, the institution's Carnegie classification (1 = doctoral, 2 = master's, 3 = baccalaureate), and the number of institutions of this type both in the overall sample and United States as a whole. These weighting variables were also calculated and analyzed for the 1,379 nonrespondents. In the *SI Appendix*, we describe detailed results for the unweighted analysis; results for the weighted analysis were comparable. (Further descriptions of methods, analyses, and results, as well as candidate summaries, full CVs, and cover letters, appear in the *SI Appendix*.) We also conducted multiple validity checks to assess the representativeness of the 34.02% sample of respondents. First, we offered \$25 to 90 solicited subjects if they provided data; 82 did so (91.1% response rate), and the distribution of these data matched the full sample. Second, experiment 4 was an additional validation, using 35 engineering faculty members who evaluated full CVs, rather than

narrative summaries of applicant credentials. Third, in experiment 5, 127 faculty members rated a single candidate from experiment 1, presented alternatively as male or female.

Experiment 1. The main experiment ($n = 363$: 182 women, 181 men) consisted of a between-subjects contest between identically qualified female and male applicants for an assistant professorship who shared academic credentials and lifestyles (plus the Y foil candidate). The six lifestyles studied were single without children, married without children, married with children and stay-at-home spouse, married with children and spouse working outside home, married with children and spouse working inside home, and divorced with children. Candidates' children were always described as two preschoolers. A random stratified sampling procedure was used (*SI Appendix*). Our data revealed an overall strong preference for female applicants over identically qualified males who shared the same lifestyle (Fig. 1). This preference for women was observed across all three Carnegie classifications of institutions, all four fields, and both genders of faculty, with the exception of male economists (see following). Effect sizes for this preference for women were large (d s between 0.8 and 1.42). Women were ranked first by 67.3% of faculty, representing a highly significant 2:1 advantage ($n = 339$; $\chi^2 = 40.38$; $P < 0.0001$). There was no evidence that women were preferred more often in some fields than others: women were strongly preferred in biology, engineering, economics, and psychology, with χ^2 s ranging from 3.89 to 19.17 and all P s < 0.05 . With the single exception of economics, there was no difference between male and female faculty in their strong preference for female applicants; in economics, male faculty rated identically qualified male and female candidates as equally hirable (54.8% for male candidates vs. 45.2% for females; $n = 31$; $\chi^2 = 0.29$; $P = 0.59$). It is worth noting that women economists preferred women candidates two to one, 68.3% to 31.7% ($n = 41$; $\chi^2 = 5.49$; $P = 0.019$), but most economics faculty members are male. Thus, men's votes carry more weight in economics hiring decisions, although men in this field were gender-neutral in their hiring preference, not antifemale.

An overall comparison of applicants within each of the six lifestyles showed the same strong preference for women with no effect of specific lifestyle (i.e., being married or single or being with or without daycare-age children did not change the highly significant 2:1 female advantage; Fig. 2; all preferences for women were significant within each lifestyle with the exception of mothers with spouses running home-based businesses). The most common lifestyle for assistant professor applicants is single without children; here women were strongly and equivalently preferred by both male and female faculty members: 66.7% and 75.9%, respectively ($n = 62$; $\chi^2 = 10.90$; $P = 0.001$; there was no difference in men's and women's preference for women: $\chi^2 = 0.63$; $P = 0.43$).

The data from experiment 1 were reanalyzed using sample weights. As reported in the *SI Appendix*, these weighted analyses reaffirmed our conclusions, suggesting that any nonrandomness costs were trivial. Considered alongside the same results, which were again obtained with the 91% response-rate paid sample (described later), this strongly suggests our results generalize to the four fields studied in the population of US colleges and universities.

Experiment 2. In real-world hiring, competing applicants do not necessarily share the same lifestyle, and bias has historically been alleged to disadvantage divorced mothers, for example, while advantaging married fathers. Experiment 2 ($n = 144$: 80 men, 64 women) consisted of a hiring comparison with targeted hypotheses, pitting applicants with differing lifestyles that may occur in the real world, counterbalanced for adjective gender. We found that the 2:1 preference for women in the main experiment changed in some of these cross-lifestyle contrasts (Fig. 3). One contrast was a divorced mother with two preschool-age children whose ex-husband does not relocate with her, pitted against a married father of two whose spouse is a stay-at-home mother.

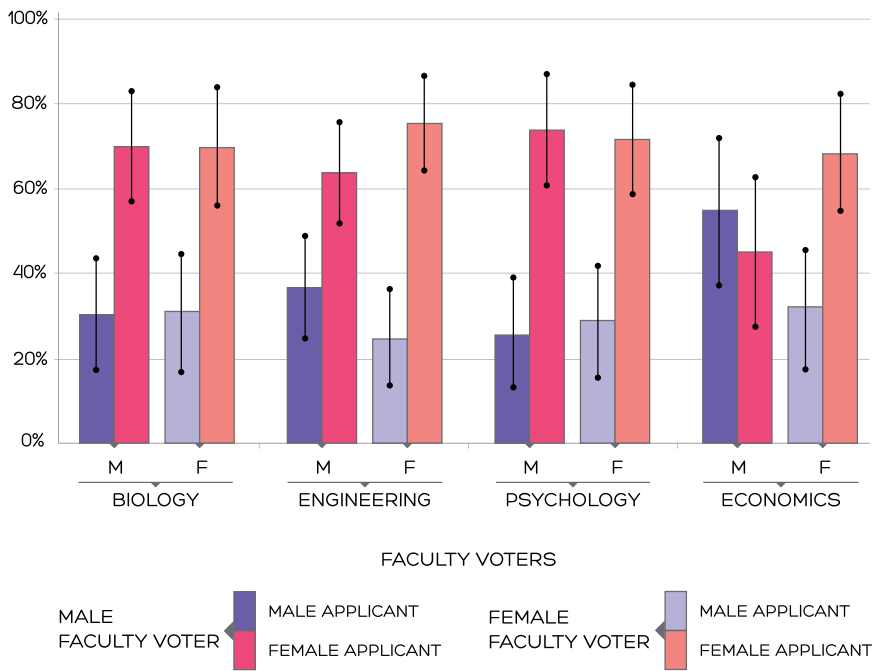


Fig. 1. Hirability of identically qualified candidates with matching lifestyles shown by field: percentage of faculty members ranking the applicant number one. Faculty members exhibit approximately a 2:1 preference for hiring women assistant professors over identically qualified men. Faculty members of both genders in all four fields expressed a strong hiring preference for female over male applicants with identical qualifications and lifestyles, compared across faculty in six counterbalanced experimental conditions ($n = 339$: 171 women and 168 men; $\chi^2 = 40.38$; $P < 0.0001$, excluding tied ranks and choice of foil), with the exception of male economists, who ranked both genders equivalently. Engineering data include validation sample of 35 engineering faculty.

Another contrast was a married father with a stay-at-home spouse competing against a single woman with no children. We found that female faculty strongly and significantly preferred divorced mothers over identically qualified married fathers (71.4% vs. 28.6%, respectively), whereas male faculty members showed the opposite but nonsignificant trend (42.9% favoring divorced mothers vs. 57.1% favoring married fathers). The overall analysis combining both genders of faculty showed a significant difference between female and male faculty's preferences for married fathers versus divorced mothers ($n = 63$; $\chi^2 = 5.14$; $P = 0.04$). Because female faculty members are underrepresented in math-intensive fields, women's strong preference for divorced mothers over married fathers may be limited in its effect on faculty hiring decisions. [Note, however, that in the experiment 1 matching-lifestyle contest between identically qualified divorced mothers and divorced fathers, male faculty members chose divorced mothers 60.7% of the time, and female faculty preferred divorced mothers 70.0% of the time (overall preference for these divorced women = 65.5%; $n = 58$; $\chi^2 = 5.59$; $P = 0.018$).] In addition, in the competition between a married father and a single woman without children, everyone preferred the single woman: almost 3:1 for male faculty (73.0%) and 4:1 for female faculty (78.1%), which are statistically equivalent ($n = 69$; $\chi^2 = 0.25$; $P = 0.62$).

Experiment 3. One lifestyle factor of current national policy interest is the effect of taking parental leave during graduate school. In experiment 3 ($n = 204$: 109 women, 95 men), we explored faculty's hiring preferences for candidates (with children) who take versus do not take 1-y parental leaves in graduate school. We contrasted two identically qualified members of one gender (all parents, with male and female candidates evaluated in separate conditions) who either took or did not take leaves, counterbalanced for adjective gender. Male and female faculty responded differently to hypothetical candidates based on candidate gender and leave status ($n = 190$; $\chi^2 = 4.21$; $P = 0.04$; Fig. 4). Male faculty members preferred 2:1 mothers who took 1-y leaves over mothers matched in academic quality who took no leaves (65.9% to 34.1%; $n = 44$; $\chi^2 = 4.45$; $P = 0.049$), but these male faculty members showed no preference between fathers who took vs. did not take leaves (48.9% vs. 51.1%). Female faculty members also showed no preference regarding fathers'

leave status (53.6% with leave vs. 46.4% with no leave). However, female faculty members ($n = 45$) showed a trend toward preferring mothers who took no extended leaves over equally qualified mothers who took leaves: 62.2% to 37.8%. Although this trend was not significant when evaluated solely within female faculty members, in an overall analysis, female and male faculty members showed significantly different preferences for mothers with versus without parental leaves ($n = 89$; $\chi^2 = 7.05$; $P = 0.01$).

Experiment 4. Experiment 4 was a validation study to determine whether rankings of candidates based on narrative summaries would be replicated if we used full CVs. The use of narrative summaries in experiments 1–3 and 5 was essential in the national cross-field data collection to avoid problems of noncomparability inherent in sending the same CV to small teaching-intensive colleges and large research-intensive universities. This is because a vita viewed as a “9.5 out of 10” at a doctoral-intensive institution would show more research productivity than is typical of

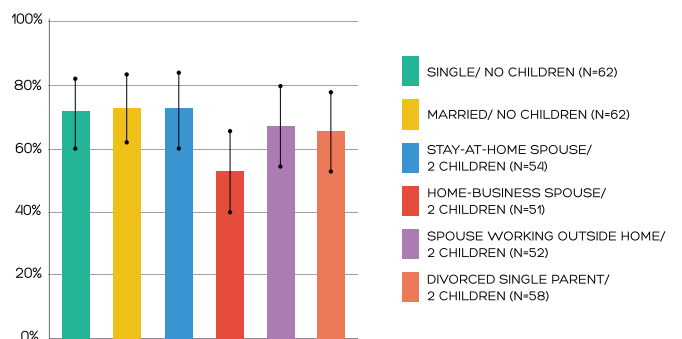


Fig. 2. Percentage of female applicants chosen over identically qualified men with matching lifestyles, shown by lifestyle. Percentage of faculty members who preferred to hire the female applicant over the identically qualified male applicant with the same lifestyle, shown for six different lifestyles ($n = 339$; all preferences for women over men are significant with the exception of that for mothers with spouses running home-based businesses, with significance levels ranging from $z = 2.23$ ($P = 0.025$) to $z = 3.18$ ($P = 0.0013$)).

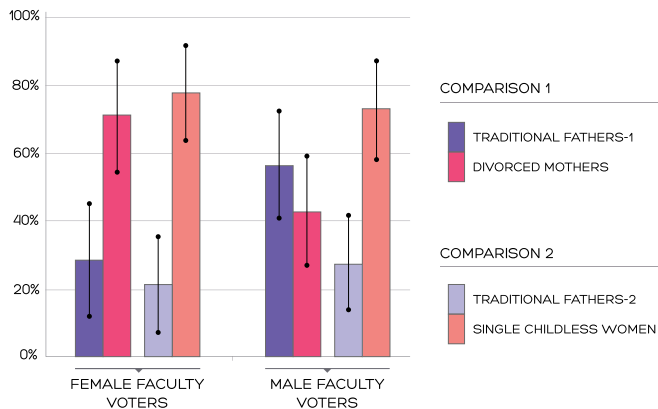


Fig. 3. Hirability of identically qualified candidates with different lifestyles: percentage of faculty members ranking the applicant number one. In a comparison between a divorced mother and an identically qualified traditional father with a stay-at-home wife (both with two preschoolers), female faculty members chose the divorced mother 71.4% of the time and the traditional father 28.6% of the time, revealing a significant preference for divorced mothers ($n = 28$; $\chi^2 = 5.14$; $P = 0.036$). In contrast, male faculty members chose the traditional father 57.1% of the time and the divorced mother 42.9% of the time ($n = 35$; $\chi^2 = 0.71$; $P = 0.50$). Male and female faculty members showed significantly different preferences for married fathers versus divorced mothers ($n = 63$; $\chi^2 = 5.14$; $P = 0.04$). In a separate condition, a comparison between single, childless women and traditional fathers showed that single, childless women are strongly preferred by both genders of faculty, independently and with both genders combined (aggregate $n = 32$ women and 37 men; total = 69; $\chi^2 = 17.75$; $P < 0.0001$).

vitae of applicants to most small colleges emphasizing teaching. In experiment 4, 35 engineering professors at 27 doctoral-intensive institutions (19 men, 16 women) ranked three applicants for whom they had full CVs; this was the same task used in experiment 1, except with full CVs substituted for narrative summaries. The engineers ranked the female significantly higher than the male applicant ($n = 33$; $\chi^2 = 8.76$; $P = 0.003$) by nearly the same margin found among engineering faculty members in experiment 1. The woman candidate was chosen over the identical man by an even larger margin than in experiment 1 (75.8%, or 25 of 33 engineers chose the woman, with two choosing the foil, vs. 66.7% of 84 engineers choosing the woman in experiment 1), although this difference was not significant ($n = 117$; $\chi^2 = 0.92$; $P = 0.34$). This finding confirms that narrative summaries were suitable proxies for CVs and resulted in equivalent preference for women, while having the advantage of comparability across institutions, fields, and subfields.

Experiment 5. Would faculty members still prefer female applicants if faculty are asked to evaluate one individual, rather than choosing among men and women? It seems possible that comparing female and male applicants could tilt faculty responses in a socially desirable direction (i.e., endorsing gender diversity). However, rating one applicant avoids socially desirable responding because there is no contrast between a man and woman. If, comparing ratings across faculty members, people rate male applicants higher than females when presented with only one applicant, this could suggest implicit antifemale biases that emerge when raters have no explicit male–female comparison. It is possible that raters may subconsciously or consciously counter such biases when making a gender comparison, as might have occurred in the earlier experiments, perhaps out of desire for gender diversity (or at least the appearance of it), when explicitly forced to choose between a man and woman. Conversely, if faculty members rate women equal to or higher than otherwise identical men when they are presented with only one applicant, this suggests they have internalized the values of gender diversity and exhibit a

desire for it, even when evaluating an applicant not pitted against an opposite-gender competitor.

We asked 127 faculty members (63 women, 64 men, from 82 institutions) to rate a single applicant, using the descriptive narrative summaries from experiment 1, on a scale of 1 to 10, ranging from “1 = cannot support” to “10 = truly extraordinary/exceptional; do whatever it takes to recruit.” Unsurprisingly, given the high level of competence of short-listed applicants for tenure-track positions, 89% of faculty rated the applicant on the upper half of the scale. A two-way between-subjects ANOVA revealed a significant main effect favoring the female applicant [$F(1,123) = 16.48$; $P < 0.0001$]; she was rated one scale point higher than an identical man (8.20 vs. 7.14; $\eta = 0.12$). There was no significant effect for gender of faculty rater [$F(1,123) = 2.75$; $P = 0.10$], but there was a marginal interaction between gender of faculty rater and gender of applicant [$F(1,123) = 3.36$; $P = 0.07$; $\eta = 0.03$], reflecting a larger down-rating of male applicants by male than by female faculty members (however, this effect was not significant after Bonferroni correction). Thus, faculty members of both genders favored the female applicant over the identically qualified male, which is consistent with the preference for women observed in the earlier experiments, in which faculty members chose among applicants of both genders. The existence of a preference for women when faculty rate only one applicant suggests that norms and values associated with gender diversity have become internalized in the population of US faculty.

General Discussion

Our experimental findings do not support omnipresent societal messages regarding the current inhospitability of the STEM professoriate for women at the point of applying for assistant professorships (4–12, 26–29). Efforts to combat formerly widespread sexism in hiring appear to have succeeded. After decades of overt and covert discrimination against women in academic hiring, our results indicate a surprisingly welcoming atmosphere today for female job candidates in STEM disciplines, by faculty of both genders, across natural and social sciences in both math-intensive and non-math-intensive fields, and across fields already well-represented by women (psychology, biology) and those still poorly represented (economics, engineering). Women struggling

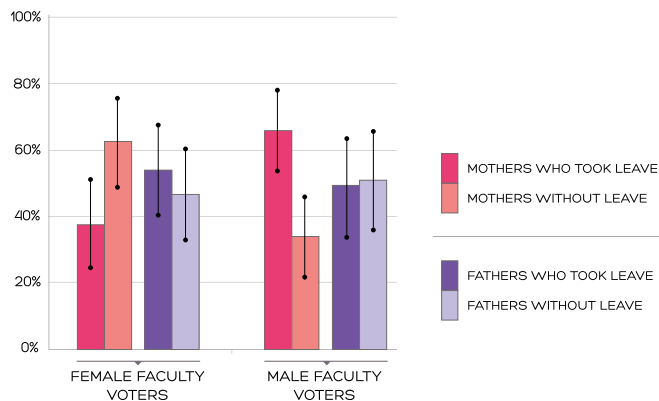


Fig. 4. Effect of 1-y parental leave on hirability: percentage of faculty members ranking the applicant number one. Male faculty members preferred mothers who took 1-y parental leaves 65.9% of the time over identically qualified mothers who did not take leaves, chosen 34.1% of the time ($n = 44$; $\chi^2 = 4.45$; $P = 0.049$). In contrast, female faculty members showed the reverse (nonsignificant) trend, choosing mothers who did not take leaves 62.2% of the time over mothers who took leaves, chosen 37.8% of the time ($n = 45$; $\chi^2 = 2.69$; $P = 0.135$). Male and female faculty members showed significantly different preferences for mothers who did versus did not take parental leaves ($n = 89$; $\chi^2 = 7.05$; $P = 0.01$). Neither female nor male faculty members exhibited a hiring preference regarding fathers’ leave status, with values ranging between 46.4% and 53.6% ($n = 56$ women and 45 men; total $n = 101$).

with the quandary of how to remain in the academy but still have extended leave time with new children, and debating having children in graduate school versus waiting until tenure, may be heartened to learn that female candidates depicted as taking 1-year parental leaves in our study were ranked higher by predominantly male voting faculties than identically qualified mothers who did not take leaves.

Our data suggest it is an auspicious time to be a talented woman launching a STEM tenure-track academic career, contrary to findings from earlier investigations alleging bias (3–13), none of which examined faculty hiring bias against female applicants in the disciplines in which women are underrepresented. Our research suggests that the mechanism resulting in women's underrepresentation today may lie more on the supply side, in women's decisions not to apply, than on the demand side, in antifemale bias in hiring. The perception that STEM fields continue to be inhospitable male bastions can become self-reinforcing by discouraging female applicants (26–29), thus contributing to continued underrepresentation, which in turn may obscure underlying attitudinal changes. Of course, faculty members may be eager to hire women, but they and their institutions may be inhospitable to women once hired. However, elsewhere we have found that female STEM professors' level of job satisfaction is comparable to males', with 87%-plus of both genders rating themselves "somewhat to very" satisfied in 2010 (figure 19 in ref. 14). Also, it is worth noting that female advantages come at a cost to men, who may be disadvantaged when competing against equally qualified women. Our society has emphasized increasing women's representation in science, and many faculty members have internalized this goal. The moral implications of women's hiring advantages are outside the scope of this article, but clearly deserve consideration.

Real-world data ratify our conclusion about female hiring advantage. Research on actual hiring shows female Ph.D.s are disproportionately less likely to apply for tenure-track positions, but if they do apply, they are more likely to be hired (16, 30–34), sometimes by a 2:1 ratio (31). These findings of female hiring advantage were especially salient in a National Research Council

report on actual hiring in six fields, five of which are mathematically intensive, at 89 doctoral-granting universities (encompassing more than 1,800 faculty hires): "once tenure-track females apply to a position, departments are on average inviting more females to interview than would be expected if gender were not a factor" (ref. 16, p. 49). [See *SI Appendix* for descriptions of other audits of actual hiring that accord with this view, some dating back to the 1980s. Many studies have argued (see ref. 14) that because only the very top women persist in math-intensive fields, their advantage in being hired is justified because they are more competent than the average male applicant. This is why an accurate evaluation of gender preference in hiring depends on data from an experiment in which competence is held constant.] Thus, real-world hiring data showing a preference for women, inherently confounded and open to multiple interpretations because of lack of controls on applicant quality, experience, and lifestyle, are consistent with our experimental findings.

Although the point of entry into the professoriate is just one step in female faculty's journey at which gender bias can occur, it is an extremely important one. Elsewhere we have examined subsequent factors in women's versus men's academic science careers, such as gender differences in remuneration, research productivity, citations, job satisfaction, and promotion and concluded that with some exceptions, the academy is gender-fair (14). We hope the discovery of an overall 2:1 preference for hiring women over otherwise identical men will help counter self-handicapping and opting-out by talented women at the point of entry to the STEM professoriate, and suggest that female underrepresentation can be addressed in part by increasing the number of women applying for tenure-track positions.

ACKNOWLEDGMENTS. We thank K. Clermont, D. Dunning, M. Macy, and J. Mendle for comments; F. Vermeylen, M. Wells, J. Bunge, F. Thoenmes, and S. Schwager for statistical advice; M. Bevia for graphic design; seven anonymous reviewers, one anonymous statistician who replicated our findings, and the editor; and 873 faculty who generously participated in our experiments. Research was supported by NIH Grant 1R01NS069792-01.

- Blau FD, Currie JM, Crosron RTA, Ginther DK (2010) Can mentoring help female assistant professors? Interim results from a randomized trial. *Am Econ Rev* 100(2):348–352.
- Handelsman J, et al. (2005) More women in science. *Science* 309(5738):1190–1191.
- Budden A, et al. (2008) Response to Webb et al.: Blind review: Accept with minor revisions. *Trends Ecol Evol* 23(7):353–354.
- Wenneras C, Wold A (1997) Nepotism and sexism in peer-review. *Nature* 387(6631):341–343.
- Bug A (2010) Swimming against the unseen tide. *PhysicsWorld*. Available at physicsworld.com/cws/article/print/2010/aug/02/swimming-against-the-unseen-tide. Accessed March 18, 2015.
- 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.
- Steinpreis R, Anders RK, Ritzke KD (1999) The impact of gender on the review of the CVs of job applicants and tenure candidates: A national empirical study. *Sex Roles* 41(7/8):509–528.
- Knobloch-Westerwick S, Glynn CJ, Hugu M (2013) The Matilda effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Sci Commun* 35(5):603–625.
- Reuben E, Sapienza P, Zingales L (2014) How stereotypes impair women's careers in science. *Proc Natl Acad Sci USA* 111(12):4403–4408.
- Sheltzer JM, Smith JC (2014) Elite male faculty in the life sciences employ fewer women. *Proc Natl Acad Sci USA* 111(28):10107–10112.
- Hill C, Corbett C, St. Rose A (2010) *Why so Few? Women in Science, Technology, Engineering, and Mathematics* (American Association of University Women, Washington, DC).
- Institute of Medicine, National Academy of Sciences, and National Academy of Engineering (2007) *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering* (The National Academies Press, Washington, DC).
- West MS, Curtiss JW (2006) *AAUP Gender Equity Indicators 2006* (American Association of University Professors, Washington, DC).
- Ceci SJ, Ginther DK, Kahn S, Williams WM (2014) Women in academic science: A changing landscape. *Psychol Sci Publ Interest* 15(3):75–141.
- Ginther DK, Kahn S (2009) Does science promote women? Evidence from Academia 1973–2001. *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, eds Freeman RB, Goroff DF (University of Chicago Press, Chicago, IL), pp 163–194.
- National Research Council (2009) *Gender Differences at Critical Transitions in the Careers of Science, Engineering and Mathematics Faculty* (National Academies Press, Washington, DC).
- Kaminski D, Geisler C (2012) Survival analysis of faculty retention in science and engineering by gender. *Science* 335(6070):864–866.
- Ceci SJ, Williams WM (2011) Understanding current causes of women's underrepresentation in science. *Proc Natl Acad Sci USA* 108(8):3157–3162.
- Davidson HK, Burke MJ (2000) Sex discrimination in simulated employment contexts: A meta-analytic investigation. *J Vocat Behav* 56(2):225–248.
- Williams WM, Ceci SJ (2012) When scientists choose motherhood. *Am Sci* 100(2):138–145.
- Mason MA, Goulden M, Frasch K (2009) Why graduate students reject the fast track. *Academe* 95(1):11–16.
- Correll SJ, Benard S, Paik I (2007) Getting a job: Is there a motherhood penalty? *Am J Sociol* 112(5):1297–1339.
- Clarkberg M, Moen P (2001) Understanding the time-squeeze: Married couples preferred and actual work-hour strategies. *Am Behav Sci* 44(7):1115–1136.
- Cuddy AJC, Fiske ST, Glick P (2004) When professional become mothers, warmth doesn't cut the ice. *J Soc Issues* 60(4):701–718.
- Diekmann AB, Eagley A (2000) Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Pers Soc Psychol Bull* 26(10):1171–1188.
- Foschi M, Lai L, Sigerson K (1994) Gender and double standards in the assessments of job candidates. *Soc Psychol Q* 57(4):326–339.
- Monroe K, Ozyurt S, Wrigley T, Alexander A (2008) Gender equality in academia: Bad news from the trenches, and some possible solutions. *Perspect Polit* 6(2):215–233.
- Urry M (2012) Why we are biased against women in science. CNN. Available at inamerica.blogs.cnn.com/2012/10/01/opinion-why-are-we-biased-against-women-in-science/. Accessed August 6, 2014.
- MIT (1999) A study of the status of women faculty in science at MIT. Available at: web.mit.edu/fnl/women/women.pdf. Accessed March 18, 2015.
- Wolfinger NH, Mason MA, Goulden M (2008) Problems in the pipeline: Gender, marriage, and fertility in the ivory tower. *J Higher Educ* 79(4):388–405.
- Glass C, Minnotte K (2010) Recruiting and hiring women in STEM fields. *J Divers High Educ* 3(4):218–229.
- Irvine AD (1996) Jack and Jill and employment equity. *Dialogue* 35(02):255–292.
- Kimura D (2002) *Preferential hiring of women*. University of British Columbia Reports. Available at: www.safs.ca/april2002/hiring.html. Accessed March 18, 2015.
- Seligman C (2001) Summary of recruitment activity for all full-time faculty at the University of Western Ontario by sex and year. Available at: www.safs.ca/april2001/recruitment.html. Accessed March 18, 2015.