



Employee Performance Management

Research: How Women Undersell Their Work

by Marc J. Lerchenmueller, Olav Sorenson and Anupam B. Jena

December 20, 2019



HBR Staff/Yagi Studio/Getty Images

Summary. Many factors contribute to the gender disparities in academia. Productivity differences, however, cannot account for them. Instead, research suggests that women receive less recognition than men for equivalent accomplishments. Just why they receive less... [more](#)

Despite a narrowing of the gender gap in science, women still lag behind men, especially at the highest levels. In the life sciences, for example, women now earn just as many doctoral degrees as men. But they hold only one out of four full professorships at research universities in the [United States](#). Women in the life

sciences also earn less and receive less research funding than men.

Many factors contribute to these gender disparities in academia. Productivity differences, however, cannot account for them. Instead, research suggests that women receive less recognition than men for equivalent accomplishments. Just *why* they receive less attention has been an open question.

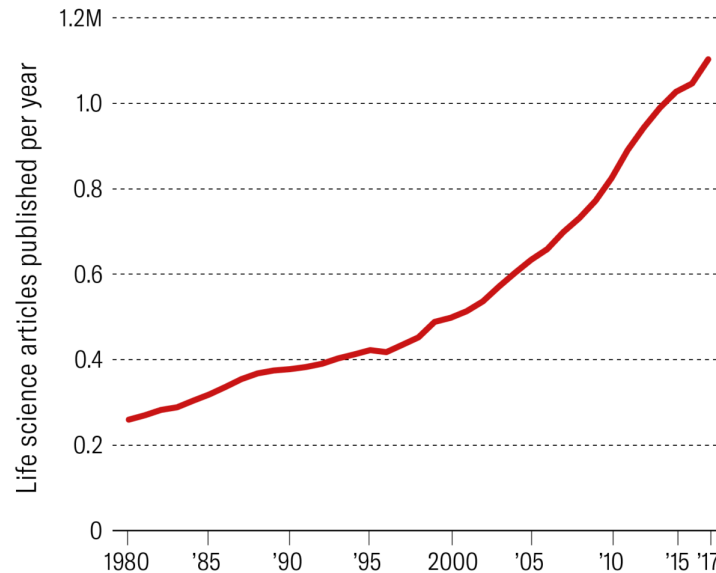
Our research examined whether women and men differ in the degree to which they promote – or spin – their accomplishments by using positive terms like “novel,” “unique,” or “unprecedented” when describing their research. In a study published in the *British Medical Journal*, we document that women use fewer of these positive adjectives in research articles. These differences in presentation, in turn, appear to influence the amount of attention that their articles receive.

The importance of self-promotion

The rate of scientific publication has exploded over the last few decades. The life sciences alone recently passed the 1 million article-per-year threshold, an extraordinary fourfold increase in scholarly output compared to the early 1980s. In fact, the roughly 20 million articles published since 1980 represent about 75% of all life sciences articles ever published and recorded.

The Rate of Scientific Publication Has Increased Exponentially

In the life sciences alone, more than one million articles are published annually — a fourfold increase in scholarly output compared to the early 1980s.



Source: Publications per year recorded in the PubMed database, the global standard bibliographic reference in the life sciences



But with this wealth of information has come a poverty of attention. Scientists have had to become ever more choosy in deciding what to read and how to allocate their time. Journal prestige undoubtedly provides a signal of research importance. But even in these selective outlets, the offerings are still overwhelming. Self-promotion has almost certainly become more important than ever in capturing scientists' attention.

Self-promotion differs in several important ways from other means of vying for scarce resources, such as applying for grants or negotiating for pay. Individuals often have substantial control over the extent to which they promote themselves. And opportunities to self-promote abound, from sharing research via social media to framing results in articles and presentations in the most favorable light.

We examined potential gender differences in self-promotion by analyzing the titles and abstracts of approximately 6.2 million research articles in the life sciences, published over a 15-year period (2002–2017). We assembled this data from two sources. Our information on articles comes from the *PubMed* database, the most complex article database for the life sciences. We determined author gender probabilistically, by using the first names of the authors and the *Genderize* database. If at least 90% of people with a particular name were women, we coded the author as a woman.

Our analyses focused on language use in article titles and abstracts, because these passages represent some of the most important text to convey the main findings. Scientists often use these short passages to determine what to read in detail. We counted the number of times authors used a set of 25 distinctively positive words — like “novel,” “unique,” “prominent,” “excellent,” and “unprecedented.” We also used natural language processing to characterize the text surrounding these positive words.

One difficulty in studying language usage stems from the fact that every article reports a unique piece of research, yet these findings differ in their novelty and in their importance. If men happen to conduct more novel research than women, then men’s more frequent use of phrases like “novel finding” may simply reflect the nature of the work, as opposed to self-promotion. To ensure a comparison of apples to apples, we only compared publications that investigated topics of similar novelty (determined from the keywords assigned to the articles). We also only compared articles published in the same journals in the same years to account for differences in journal prestige and subject area.

A second difficulty stems from the question of attribution. Science, and especially the life sciences, has become a team sport.

Determining who wrote what in an article with three, four, or more authors can prove difficult. Here, norms help us. Long-standing tradition in the life sciences usually assigns the first author position to the junior scientist who executed the project and the last author position to the senior investigator who funded and who often conceived of the project. We therefore compared articles written by teams of female first and last authors to teams that involved at least one man in either lead position.

The gender gap in self-promotion

We found that articles written by female first and last authors were up to 21% *less likely* to use positive terms to frame their research findings than comparable articles published in the most prestigious journals with at least one man in a lead author position. Additional analyses confirmed that these positive words typically qualified the findings: we would frequently see word combinations such as “novel approach,” “unique mechanism,” or “promising result.” Both men and women used the positive words to describe their findings, but women used them less frequently.

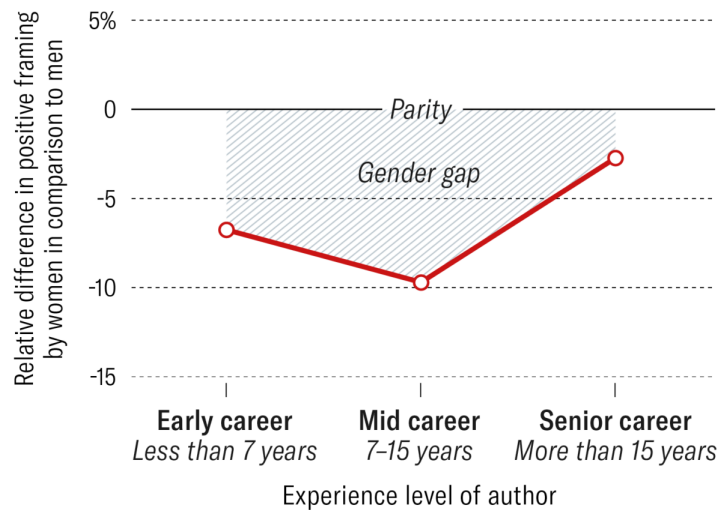
We also examined whether these gender differences in self-promotion had consequences by determining whether they influenced the number of downstream citations (when the work is referenced by others). We found that articles with positive words received up to 13% more citations compared to research of similar novelty published in the same journals but without the positive framing. This apparent advantage of positive framing was largest for articles published in the most influential journals, those with a journal impact factor exceeding 10. In other words, authors that did not self-promote paid a price – in terms of receiving less attention – particularly when they published in the most prominent journals.

We also explored whether gender differences in self-promotion varied as a function of career stage. In the **top panel of the figure below**, you can see the difference in promotion words between male and female lead-authored articles by career stage of the first author. (We controlled for the experience of the last author as well as for potential cohort effects among the first authors by including a control for when these authors first published, to account for changes in language use over time). The **bottom panel of the figure** depicts the parallel results for the career stages of the last author.

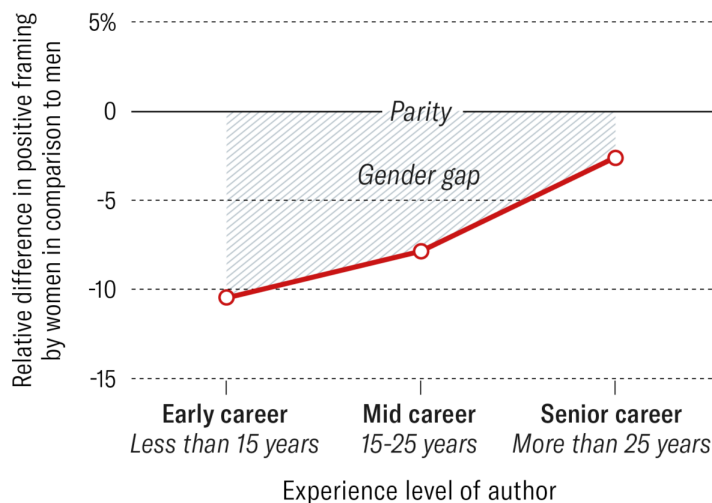
The Gender Gap in Positive Framing of Research Closes as Women Rise Through the Ranks

Traditionally, the two lead authors in life science research are listed first and last in papers about the studies. For both first- and last-listed authors, gender differences in self-promotion appeared most pronounced at the early- and mid-career stages.

First-listed authors (typically a junior scientist who executed the research project):



Last-listed authors (typically a senior investigator who funded — and often conceived — of the project):



Source: Publication trajectories associated with individual author IDs from the PubMed and Author-ity databases (1985-2009)



For both first and last authors, gender differences in self-promotion appeared most pronounced at early- and mid-career

stages. As women rose through the ranks, their usage of positive words increased. At the most senior levels, the disparity disappears (the 95% confidence intervals — not shown here — straddle the parity line). Some of this effect may stem from selection, with the more self-promotional authors having better odds of reaching these senior ranks. But women may also present their research more confidently as they gain seniority.

Does this mean women should hype their research more? Our research cannot inform the optimal degree of positive framing for the scientific community. Generally speaking, the language used should accurately reflect the quality and importance of the findings. Our study reveals that women self-promote less than men, but we cannot say whether women understate their achievements or whether men overstate them.

Although our research has focused on the life sciences, we suspect that these gender disparities in self-promotion occur in a wide variety of settings, probably contributing to the societal gender gaps in pay and promotion. So it seems fair to say that women would do well to promote their accomplishments more. But the onus does not reside only with them. Male colleagues also need to encourage women and ensure that they do not get penalized for the self-promotion they do.



Marc J. Lerchenmueller is assistant professor of technological innovation and management science at the University of Mannheim, and a research fellow at Yale University. He received his bachelor's degree from Imperial College London, master's degrees from the University of Oxford and Yale University, and his doctorate degree from WHU.



Olav Sorenson is the Frederick Frank '54 and Mary C. Tanner Professor of Management, Director of the Core Curriculum, Co-Director of the Initiative on Leadership and Organization, and Professor of Sociology (by courtesy), at Yale University. He has previously held positions at London Business School, UCLA, the University of Chicago, and the University of Toronto. He received his bachelor's degree from Harvard College and his master's and doctoral degrees from Stanford University.



Anupam B. Jena is the Ruth L. Newhouse associate professor of health care policy at Harvard Medical School, an internist at Massachusetts General Hospital, and a faculty research fellow at the National Bureau of Economic Research. He received his bachelor's degree from MIT and his MD/ PhD degrees from the University of Chicago.



Read more on **Employee performance management** or related topics
Gender, Personal productivity and **Education institutions**

Representation and Salary Gaps by Race-Ethnicity and Gender at Selective Public Universities

- October 2017
- [Educational Researcher](#) 46(7):343-354
- DOI:[10.3102/0013189X17726535](#)
- **Authors:** Diyi Li and Cory Koedel

https://www.researchgate.net/publication/320047431_Representation_and_Salary_Gaps_by_Race-Ethnicity_and_Gender_at_Selective_Public_Universities

Li, Diyi & Koedel, Cory. (2017). Representation and Salary Gaps by Race-Ethnicity and Gender at Selective Public Universities. *Educational Researcher*. 46. 343-354. 10.3102/0013189X17726535. We use data from 2015–2016 to document faculty representation and wage gaps by race-ethnicity and gender in six fields at selective public universities. Consistent with widely available information, Black, Hispanic, and female professors are underrepresented and White and Asian professors are overrepresented in our data. Disadvantaged minority and female underrepresentation is driven predominantly by underrepresentation in science and math intensive fields. A comparison of senior and junior faculty suggests a trend toward greater diversity, especially in science and math intensive fields, because younger faculty are more diverse. However, Black faculty are an exception. We decompose racial-ethnic and gender wage gaps and show that academic field, experience, and research productivity account for most or all of the gaps. We find no evidence of wage premiums for individuals who improve diversity, although for Black faculty we cannot rule out a modest premium.

INSTITUTIONAL DETERMINANTS OF FACULTY DIVERSITY

Authors: Baniyelme D. Zoogah and Susan Josephs

Published Online: 30 Nov 2017

<https://doi.org/10.5465/ambpp.2005.18778663>

<https://journals.aom.org/doi/10.5465/ambpp.2005.18778663>

Abstract

Research on the under-representation of minority faculty in academia (faculty diversity) has centered on two schools of thought. The 'academic pipeline' argument posits that the under-representation of minority faculty results from low supply of minority doctoral students. The racial discrimination argument, on the other hand, argues that the historical underpinnings of racism persist and stymie minority doctoral students even if supply of minorities was high. We advance a third view -- institutional argument -- as a third cause of minority faculty under-representation. Using Oliver's (1991) institutional framework, this study finds that contextual, constituent, control, and content factors affect minority faculty representation in academia.

IZA DP No. 9199

Does the Gender Composition of Scientific Committees Matter?

Manuel Bagues
Mauro Sylos-Labini
Natalia Zinovyeva

July 2015

Does the Gender Composition of Scientific Committees Matter?

Manuel Bagues

Aalto University and IZA

Mauro Sylos-Labini

University of Pisa

Natalia Zinovyeva

Aalto University

Discussion Paper No. 9199

July 2015

IZA

P.O. Box 7240

53072 Bonn

Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Does the Gender Composition of Scientific Committees Matter?*

An increasing number of countries are introducing gender quotas in scientific committees. We analyze how a larger presence of female evaluators affects committee decision-making using information on 100,000 applications to associate and full professorships in all academic disciplines in two countries, Italy and Spain. These applications were assessed by 8,000 evaluators who were selected through a random draw. A larger number of women in evaluation committees does not increase either the quantity or the quality of female candidates who qualify. If anything, when evaluators' are not familiar with candidates' research area, gender-mixed committees tend to be less favorable towards female candidates than all-male committees, with the exception of evaluations to full professorships in Spain. Data from 300,000 individual voting reports suggests that men become less favorable towards female candidates as soon as a woman joins the committee.

JEL Classification: J71, J16

Keywords: scientific committees, gender discrimination, randomized natural experiment

Corresponding author:

Manuel Bagues
Aalto University
Department of Economics
P.O. Box 21240
00076 Helsinki
Finland
E-mail: manuel.bagues@aalto.fi

* This paper combines information that was previously reported in two working papers: "Do Gender Quotas Pass the Test? Evidence from Academic Evaluations in Italy" and "Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment." We would like to thank Olympia Bover, Irma Clots-Figueras, Sara de la Rica, Gemma Derrick, Juan Jose Dolado, David Dorn, Silvia Dorn, Berta Esteve-Volart, Luis Garicano, Marco Giarratana, Elena Martínez, Nic Morgan, Javier Ruiz-Castillo and participants in numerous presentations for their useful comments. We also acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada and the Spanish Ministry of Science and Technology (research grants ECO2008-06395-C05-05 and ECO2008-01116). All remaining errors are our own.

1 Introduction

The underrepresentation of women in academia remains a cause for concern among universities and policy makers around the world. In Europe, women account for 46% of PhD graduates, 37% of associate professors and only a mere 20% of full professors (European Commission 2013). Similar patterns may be observed in the US and the gender imbalance is even larger in Japan (National Research Council 2009, Abe 2012).

Several explanations may account for the lack of women in high-level positions. According to the pipeline theory, once women have entered the lower rungs of the academic career it is mainly a matter of time that they would move their way through a metaphorical pipeline to reach high-level jobs. However, in most disciplines, the share of women among faculty members remains low even after decades of improved recruitment of women at the undergraduate and the doctoral level (Ginther and Kahn 2004, 2009). Gender differences in promotion rates might also reflect differences in productivity, perhaps due to the existence of gendered roles at the household level or the lack of female mentors and role models (Blau, Curie, Croson and Ginther 2010). Some women may also devote excessive time to tasks that are socially desirable but which are not taken into account in promotion decisions (Vesterlund, Babcock and Weingart 2014). Furthermore, some authors have pointed out that women are less likely to apply for promotions (Bosquet, Combes and Garcia-Peñalosa 2013; De Paola, Ponzio and Scoppa 2015), perhaps due to the existence of gender differences in the preference for competitive environments (Niederle and Vesterlund 2007; Buser, Niederle and Oosterbeek 2014) or in bargaining abilities in the labour market (Babcock, Gelfand, Small and Stayn 2006; Blackaby, Booth and Frank 2005).

Beyond these supply-side explanations, the slow progress made by women has been sometimes attributed to the existence of gender discrimination by the (mostly male) evaluators who decide on hiring and promotions.¹ In this paper we examine whether

¹Gender discrimination in academia remains a controversial issue. According to a meta-analysis by Ceci and Williams (2011), the more recent empirical evidence fails to support assertions of discrimination in manuscript reviewing, interviewing, and hiring. However, other studies show that female researchers might still receive lower evaluations than male researchers with identical characteristics (Steinpreis, Anders and Ritzke 1999, Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman

having more women in scientific committees might help to increase the chances of success of female candidates. There are several reasons for considering this hypothesis. First, there is evidence of gender segregation across different scientific subfields (Dolado, Felgueroso and Almunia 2012, Hale and Regev 2014). If men and women tend to do research in different subfields and evaluators overrate the importance of their own types of research, the lack of female evaluators might be detrimental for female candidates (Bagues and Perez-Villadoniga 2012, 2013). Second, research networks tend to be gendered (Boschini and Sjögren 2007, Hilmer and Hilmer 2007). If evaluators are mostly male, male candidates might have a better chance to be acquainted with committee members and could perhaps benefit from these connections (Zinovyeva and Bagues 2015; Bagues, Sylos-Labini, Zinovyeva 2015). Third, men may be subject to gender stereotypes or they may be biased against women reaching high-level positions. For instance, according to the World Value Survey, around 25% of US males believe that men make better political leaders and 16% think than men make better business executives.² Women are half as likely to hold such views. A similar pattern is observed in Europe.³ According to some experts in gender studies, these biases may have reached the academic world.⁴ Finally, the presence of women in evaluation committees might also improve the quality of the evaluation. It has been argued that group performance is positively correlated with the proportion of women in a group (Woolley, Chabris,

2012).

²World Value Survey Wave 6: 2010-2014. Official aggregate v.20140429. World Values Survey Association (www.worldvaluessurvey.org).

³The 6th wave of the World Value Survey provides information on eight members of the European Union: Estonia, Germany, Netherlands, Poland, Romania, Slovenia, Spain and Sweden. In these countries 30% of men and 21% women believe that men make better political leaders than women do, and 29% of men and 17% women think that men make better business executives than women do.

⁴For instance, in a report commissioned by the European Commission, the expert group *Women In Research Decision Making* argues that “(a)t the very least, having male only committees risks replicating stereotypes and bias, both regarding applicants and issues in research. Therefore, measures are needed to have both women and men fairly represented on decision making bodies” (European Commission 2008, page 27). In an another expert report on the situation of women researchers in Spain it is also argued that “there are prejudices about women among those who co-opt, promote or have the key to promotion. The bodies which control this are mostly male and, even if they are not totally conscious of it, they see an academic woman first as a woman and secondly as a colleague.” (Fundación Española para la Ciencia y la Tecnología, *Mujer y Ciencia: La situación de las Mujeres Investigadoras en el Sistema Español de Ciencia y Tecnología*, (FECYT, 2005), page 48). Other researchers have voiced similar views (Bagilhole 2005, Barres 2006, European Commission 2013, Smith et al. 2015).

Pentland, Hashmi and Malone 2010).

These arguments have reached policy makers and a number of countries have introduced quotas requiring the presence of at least 40% of women (and men) in scientific committees.⁵ Many universities and institutions all around the world also have internal guidelines requiring the presence of women in committees. But despite the increasing popularity of gender quotas in scientific committees, there are some important concerns about their effectivity (Vernos 2013). Quotas are costly for senior women, as they increase disproportionately the amount of time that female professors have to devote to evaluation committees. Furthermore, a larger presence of women in committees may not necessarily benefit female candidates. Both men and women have developed their careers in an academic environment dominated by men, and both genders may tend to associate important academic positions, and the features they require, with men, not with women (Mendez and Busenbark 2012). Women who made it to the upper echelons of academia may also exercise a type of “queen bee syndrome” in which they make exaggerated demands upon junior female staff (Staines, Jayaratne and Tavis 1973), and even if women are relatively more sympathetic towards female candidates, they may not have equal levels of voice and authority in deliberation processes (Karpowitz, Mendelberg and Shaker 2012, Brescoll 2011).

The empirical evidence on the impact of scientific committees’ gender composition has been so far scarce, typically based on small samples and rather inconclusive. Sometimes researchers seem to benefit from the presence of evaluators who share the same gender (Casadevall and Handelsman 2013, De Paola and Scoppa 2014), sometimes applicants seem to obtain relatively better evaluations from opposite-sex evaluators (Broder 1993; Ellemers, Heuvel, de Gilder, Maass and Bonvini 2004), and in some other cases gender does not seem to play any (statistically) significant role (Abrevaya and Hamermesh 2012; Jayasinghe, Marsh and Bond 2003; Moss-Racusin, Dovidio, Brescoll,

⁵In 1995 gender quotas were introduced in Finland through the amendment of the Finnish Act on Equality between Women and Men. In 1999, the European Commission stated the aim to achieve at least a 40% representation of women in Marie Curie scholarships, advisory groups, assessment panels and monitoring panels (COM(1999) 76 Final). In 2007, gender quotas were introduced in Spain within the Equality Law. More recently, France has also introduced quotas in all scientific committees (decree n 2014-997, September 2 2014, article 8).

Graham and Handelsman 2012; Steinpreis, Anders and Ritzke 1999; Williams and Ceci 2015). A brief summary of these papers is available in Table A1 in the Appendix.⁶ It is unclear whether these mixed findings reflect the idiosyncrasies of the different situations and samples analyzed in each study, or simple random sampling variation. From a policy perspective, the lack of more extensive and clear evidence is disappointing.

In this paper we analyze the role of evaluators' gender in academic evaluations using the exceptional evidence provided by two large-scale randomized natural experiments in two different countries, Spain and Italy. In order to be promoted at the university level to an associate or full professorship, in both countries researchers are required to first obtain a qualification granted by a centralized committee at the national level. These qualification exams are performed periodically in all disciplines. Our database includes information on approximately 300,000 individual evaluation reports, 100,000 applications and 8,000 evaluators in 200 different disciplines. Approximately one third of evaluation committees include no women and very rarely we observe a female majority in the committee.

The Spanish and Italian institutional arrangements provide some exceptional features. First, evaluators are selected from of a pool of eligible professors using a random draw. This allows us to consistently estimate the causal effect of committees' gender composition on evaluations. Second, evaluations are performed in all disciplines and at two different levels, qualification for assistant and full professorships. Therefore we can study how the role of evaluators' gender varies according to a number of relevant dimensions. Third, we observe extensive and detailed information on evaluators' and candidates' academic connections and their field of specialization. We use this information to test explicitly the role of gendered networks, gender segregation across different subfields, and gender stereotypes. Fourth, in the Italian case, we observe the evaluation reports submitted by individual committee members. Using this information we explore the interactions that arise within committees and we study whether the presence of women in the committee affects the voting behavior of male evaluators.

⁶A related literature also analyses the role of evaluators' gender in non academic occupations (e.g. Bagues and Esteve-Volart 2010, Booth and Leigh 2010) or in sport activities (e.g. Sandberg 2014).

Finally, we can observe the research output of candidates before the evaluation and, in the case of Spain, also their research output during the five following years. We use this information to explore whether committees with more women are better able to identify candidates with more potential.

There exist also a number of interesting institutional differences between the evaluation processes in the two countries. In Spain, evaluations involve oral presentations by the candidates, while in Italy evaluations are based only on candidates' CVs and publications. In Spain qualification leads almost automatically to promotion, while in Italy the chances are much lower. Finally, the Italian system is relatively more transparent and exposed to public scrutiny. Having data for the two slightly different institutional arrangements allows us to cross-validate the findings and to explore their robustness.⁷

We find no empirical support, either from the average in the two countries or from the majority of subsamples analyzed, to suggest that a larger presence of female evaluators in the evaluation committees has a statistically or economically significant positive effect on the chances of success of female candidates. In the case of Italy we can discard any positive impact. In fact, in the Italian sample gender-mixed committees exhibit a small but significant bias against female candidates, relative to committees composed only by male evaluators. In the Spanish case, we can reject any sizable impact. An additional woman in a committee of seven members may increase the number of women promoted by at most 0.6 p.p. (5%) or it might also decrease it by up to 1 p.p. (9%). We also examine whether committees with a relatively larger proportion of women promote better candidates, using as a proxy of candidates' quality their research output before the evaluation and during the following five years. We do not observe any significant difference in the observable quality of candidates who have qualified in committees with different gender compositions.

To get a better understanding of these findings and also to determine their validity in other contexts, we explore the different specific mechanisms suggested by the theory.

⁷For instance, some authors have argued that the degree of transparency in an evaluation procedure can affect gender biases (van den Brink, Benschop and Jansen 2010)

As discussed above, one common argument is the existence of gender segregation across academic subfields. Within each discipline, some subfields are significantly more feminized than others, but the level of segregation that we observe in our data is relatively small (2-4%). We also study segregation across research networks. Researchers are significantly more likely to have an advisor, a colleague or a co-author of the same gender (10-20%). However, although committee members tend to favor connected candidates, in these national examinations the likelihood of having a connection in a committee is low and gender segregation across research networks has only a limited effect on the evaluation outcomes. We also examine separately evaluations for high-level positions. Male evaluators might have prejudices against women being promoted to full professorships, but not to positions at lower levels of the career ladder. Results are mixed: we find support for this hypothesis in the case of Spain, but not in the case of Italy.

We also study gender stereotypes. Stereotypes may be more relevant when evaluators cannot observe accurately the quality of candidates, for instance, because evaluators and candidates are specialized in different subfields of research. In our data, when the evaluators and the candidate belong to the same subfield or research, gender does not seem to play any role. However, when information asymmetries are expected to be larger because the candidate belongs to a different subfield, female candidates have relatively better chances of success in all-male committees than in gender-mixed committees. Evidence from individual voting reports suggests that this is partly due to the impact of women in the committee on male evaluators' voting behavior. The presence of female evaluators in committees makes male evaluators tougher upon female candidates, although this effect is only marginally significant. It is unclear whether this effect reflects the information that female evaluators provide to their male colleagues or, perhaps, male evaluators' reaction to the presence of women. On the other hand, in mixed-gender committees female evaluators are slightly more favorable towards female candidates than male evaluators but this difference is not statistically significant.

2 Institutional background

Several European countries have national systems of quality assurance which are meant to guarantee the academic quality of professors in public universities. The evidence presented in this paper is based on an analysis of two variants of such systems: the Italian system known as *Abilitazione Scientifica Nazionale*, which was introduced in 2012, and the Spanish system known as *Habilitación*, which was in place between 2002 and 2006.

Both systems require candidates for associate and full professorships to qualify in national evaluations held by an academic board in the appropriate discipline. In each country, there are nearly two hundred legally defined academic disciplines, each corresponding to a certain area of knowledge. Successful candidates can then apply for a position at a given university.

In both countries, the time line of evaluations has the following steps. First, a call for applicants is announced in which candidates can apply for multiple disciplines and positions. Once the list of initial applicants is settled, committee members are randomly selected from the list of eligible evaluators in the corresponding discipline. Once the committees are formed, the evaluation process begins and once this is over, the evaluation results are made public. Rostered evaluators can potentially resign at any point of the process, something that happens in 2% of cases in Spain and in 8% of cases in Italy. Resigned evaluators are substituted by randomly selected evaluators. The procedure has also distinctive features specific to each country. We summarize these features below (see also Table A2 in the Appendix).

2.1 *Abilitazione Scientifica Nazionale*

In Italy, four out of five committee members are selected through a random draw from the pool of ‘Italian’ eligible evaluators and the remaining evaluator is drawn out of the pool of ‘foreign’ eligible evaluators. The former pool consists of full professors affiliated to Italian universities who volunteered to be members. The latter pool consists of professors affiliated to universities from OECD countries, who also volunteered

to participate in Italian evaluations. The randomization procedure is subject to one important constraint: no university can have more than one evaluator within a single committee

The eligibility of evaluators is decided in the following way. In science, technology, engineering, mathematics and medicine (STEMM), evaluators are required to have a research output above the median for full professors in the discipline in at least two of the following three dimensions: (i) the number of articles published in scientific journals, (ii) the number of citations, (iii) and the H-index. In the social sciences and the humanities (SSH), the research performance of evaluators has to be above the median in at least one of the following three dimensions: (i) the number of articles published in high quality scientific journals (in what follows, A-journals),⁸ (ii) the overall number of articles published in any scientific journals and book chapters, and (iii) the number of published books. ‘Foreign’ eligible evaluators have to satisfy the same research requirements as ‘Italian’ ones. While ‘Italian’ evaluators work *pro bono*, OECD evaluators receive €16,000 for their participation.

Evaluations are based solely on the material provided in candidates’ application packages consisting of CVs and recent publications. Committees have full autonomy regarding the criteria to be used in the evaluation and the number of qualifications to be granted. Each evaluation committee is required to draft and publish online a document describing the general criteria to be used in providing a positive assessment. Candidates may withdraw their application up until two weeks after evaluation criteria are publicized. A positive assessment of the candidate requires a qualified majority of four out of five votes. Once granted, qualifications are only valid for four years, while a negative evaluation means that candidates are excluded from participating in further national evaluations during the following two years.

An important feature of the Italian system is its extreme transparency: all the relevant information – including candidates’ and evaluators’ CVs, as well as individual evaluation reports – is published online. An independent evaluation agency appointed

⁸An evaluation agency determined with the help of several scientific committees the set of journals to be considered as high quality in each field.

by the ministry also collects and publicizes information on the research output of final candidates in the ten years preceding the evaluation, as measured by the three bibliometric indicators described above. The evaluation agency compared the research productivity of candidates in each of these three dimensions with the research productivity of professors in the category to which they applied, and committees were asked to take this information into consideration.

2.2 Habilitación

In Spain, committees are composed of seven members. In evaluations for full professorships, all evaluators are full professors based in Spanish universities or research institutes. In evaluations for associate professorships, three committee members are full professors and four evaluators are associate professors. No more than one non-university researcher is allowed to be selected as a member of the committee for a given exam. Similarly, no more than one emeritus professor may be selected as a member of a given committee.

In order to be eligible, evaluators are required to satisfy some minimum research level which is assessed by the Spanish education authority.⁹ This requirement is satisfied by approximately 81% of full professors and 70% of associate professors. Unlike the Italian system, where participation is voluntary, in Spain all eligible professors can be selected to serve in committees.

Candidates for evaluation are required to make several oral presentations in front of a committee. For candidates to full professorships, these exams have two qualifying stages. In the first stage, each candidate presents their CV and then, in the second, an example of their research work. Exams for the position of associate professor, in addition to these two stages, have an intermediate stage where candidates give a lecture on a topic randomly chosen from a syllabus proposed by the candidate. In each stage

⁹The Spanish education authority determines professors' eligibility according to the number of *sexenios* completed. *Sexenios* are granted periodically by the ministry on the basis of applicants' research output in any non-interrupted period of a maximum of six years. Eligible associate professors are required to have held at least one *sexenio* while eligible full professors are required to have held at least two *sexenios*.

evaluations are made on a majority basis. Qualifications have unlimited validity once they have been granted. The number of qualifications conceded at the national level is very limited and being accredited is, in most cases, equivalent to being promoted.

3 Data

We use data on all evaluations from the first edition of the Italian *Abilitazione Scientifica Nazionale* and on all evaluations from the Spanish *Habilitación*. In Italy, the data includes information on 184 committees, one per each academic discipline. Each committee assessed both applications to associate and to full professorships. In Spain, there are in total 967 committees in 174 disciplines, of which 502 are committees evaluating candidates for full professorships and 465 evaluate candidates for associate professorships.

The dataset includes information on eligible and actually selected evaluators, applicants, and the final outcome of the evaluation. In addition to demographic characteristics and a number of productivity measures, we have also gathered information on research networks and research specialization. In Appendix A we provide detailed information on how this information was collected, and how each variable was constructed. Below we briefly summarize the main features of the dataset.

3.1 Evaluators

In Italy, 39% of Italian female full professors and 41% of Italian male full professors volunteered and were considered eligible to sit in evaluation committees. The list of eligible evaluators includes 5,876 professors based in Italian universities and 1,365 evaluators based in OECD universities. In the average field, the pool of eligible evaluators includes 32 ‘Italian’ professors and eight ‘foreign’ professors. While approximately 20% of ‘Italian’ evaluators are women, the ‘foreign’ pool is less feminized and only 12% of ‘foreign’ evaluators are women. Taking into account the composition of both pools,

the expected share of women in the committee is around 18%.¹⁰ Approximately one out of every thirteen evaluators resigned and was replaced by another eligible evaluator. These replacements slightly increased the share of women in committees, but the difference is not statistically significant. 41% of committees include no women at all, in 35% of committees there is one woman, and only 8% of committees have a majority of female evaluators.

The CVs of potential evaluators are publicized by the Ministry. Table 1 provides descriptive statistics on the research productivity of evaluators, based on their CVs.¹¹ Eligible professors have on average 13 years of tenure in the position. They list on average 131 publications in their CVs, of which just over half are articles in scientific journals, and the rest are books, book chapters, publications in conference proceedings, patents, etc. To assess the quality of research output, in STEMM disciplines we compute the total Article Influence Score; in SSH disciplines we measure the number of articles in high impact journals, or A-journals.¹² In columns 2-4 we compare the length of tenure and the research output of male and female evaluators. For this comparison, we normalize all variables at the discipline level. Female evaluators have significantly shorter tenure than their male counterparts and they also have lower research output in almost all dimensions.

In Spain, the lists of eligible evaluators include 49,199 full professors and 61,052 associate professors.¹³ Women constitute 35% of eligible associate professors, but only 14% of full professors are women. Overall, 32% of committees are composed by only male evaluators, 39% of committees have one woman on board, while only 6% have more women than men.

¹⁰We have calculated the expected gender composition of committees using a simulation with 1,000 draws, taking into account that the lottery that decided committee composition was subject to the constraint that committees cannot include more than one member from the same university.

¹¹CVs of ‘foreign’ evaluators are not in a standardized format and are often incomplete, so they are not considered in this paper.

¹²Article Influence Score is available for all journals in the Thomson Reuters Web of Knowledge. It is related to *Impact Factor*, but it takes into account the quality of the citing journals, the propensity to cite across journals and it excludes self-citations.

¹³The Spanish data covers information from several evaluation waves, so many professors appear in the lists several times. In total, there 7,963 individual full professors and 21,979 individual associate professors in these lists.

We collect from several sources information on the research outcomes of Spanish researchers. We observe their publications in international journals covered by the Web of Knowledge and their articles and books in the Spanish language included in the database Dialnet, as well as patents in the European Patent Office in which they are listed as inventors. We also have information on researchers' activity as Ph.D. advisors and as members of dissertation committees. We compare female and male eligible evaluators, normalizing their characteristics at the level of exam and category. The results are very similar to the ones observed for the Italian academia (see columns 6-8 and 10-12 of Table 1). Female eligible evaluators have shorter tenure, and on average they published less than male researchers in the same discipline and rank. They have also lower accumulated quality-adjusted scientific production: women in STEMM disciplines have a relatively lower total Article Influence Score, and women in SSH disciplines have fewer A-journal articles.¹⁴ Among both full and associate professors, women tend to have participated less in advising and evaluating doctoral students.

3.2 Candidates

There were 69,020 applications in Italy. On average, there were 375 applications per field, with 117 of them participating in evaluations for full professor positions and 258 participating in evaluations for associate professor positions. Some candidates applied to more than one position: the average candidate participated in 1.5 evaluations.

As shown in the upper panel of Table 2, 31% of applications for the position of full professor and 41% of applications for the position of associate professor were submitted by women. Candidates for a full professorship are about 49 years old and candidates for an associate professorship are six years younger. Most applicants are based in Italian universities. About a half of applicants for associate professorships hold a permanent contract and about three fourths of applicants for full professorships do. Candidates mainly apply for an evaluation in the field in which they currently hold a permanent

¹⁴In Spain, we define A-journals following the journal rank developed by Dialnet, which categorizes journals in four groups according to their prestige.

contract.

Female applicants tend to be younger among applicants for associate professorships, and they are of a similar age to their male counterparts in evaluations for full professor (columns 3-5 and 8-10). In both cases the publication record of female candidates is significantly weaker. The only dimension in which women seem to be achieving better results than men is in publishing conference proceedings. In addition to information on productivity coming from candidates' CVs, we observe the order in which candidates submitted their applications. In principle, the timing of the application might reflect both candidates' self-confidence and quality. We normalize this variable uniformly between 0 and 1. We observe that female candidates for the post of full professor apply a bit later than their male counterparts, but no similar gender difference can be observed among candidates for associate professor positions.

In Italy, approximately 14% of applicants (9,870 out of 69,020 applications) withdrew their application once the identity and the criteria of evaluators were made public. Withdrawals were more common among female applicants. Overall, approximately 38% of applications by male candidates and 35% of applications by female candidates were successful.

As explained above, the evaluation agency of the Ministry of Education published detailed information regarding the research production of the final set of applicants in the 10 previous years. Around 38% of candidates were above the median in each of the three corresponding bibliometric dimensions. Performance according to these indicators is strongly correlated with candidates' success. Among those candidates whose quality was below the median in every dimension there was a success rate of only 4%, while among those who excelled in every dimension there was a success rate of 63%.

In addition to the final decision of the committee, we also collected information on the individual evaluation reports.¹⁵ 44% of these reports were favorable to the

¹⁵We conducted a text analysis of the individual evaluation reports. We identified approximately 9,000 different sentences that indicate the evaluator's decision to fail or to pass a given candidate. These sentences were used in approximately 279,000 of the 295,000 available individual evaluation reports. Due to the data collection problem, we are missing information on individual evaluations for

candidate and most of the time (in 83% of the cases) decisions were taken unanimously.

In Spain, overall there were 13,444 applications for full professorships and 17,799 applications for associate professorships (lower panel of Table 2). The gender ratios among applicants are very similar to the ones in Italy: around 27% of applicants to full professor are women and there are around 40% of women among applicants to associate professor. Once again, male applicants seem to have stronger research records than their female counterparts. They also tend to be slightly more successful in evaluations.

Finally, for the candidates who qualified in Spain, we collected information on their individual research productivity in a five-year period following the national evaluations and on their performance in future evaluations for promotion to full professor. This information allows us to assess the quality of selection not only in terms of candidate characteristics easily observable at the moment of the exam, but also in terms of dimensions that are difficult to observe but that are nevertheless important determinants of future productivity of the candidate.

3.3 Connections

We identify professional links between candidates and eligible evaluators. We consider all the possible interactions within each discipline, around 2.5 million possible pairs in Italy and 5 million in Spain. As shown in Table 3, the probability that a candidate and an eligible evaluator are affiliated to the same institution is around 3% in Italy and 5% in Spain. The probability that they have co-authored a paper is smaller: 1.4% in Italy and 0.4% in Spain.

In the case of Spain, we also observe if there was a student-advisor relationship or if the candidate and the eligible evaluator have participated in the same thesis committee.¹⁶ These links are relatively rare: in 0.2% of the cases the eligible evaluator is the PhD thesis director of the candidate and in 1.3% they have participated in the same thesis committee.

202 candidates.

¹⁶We consider three possible interactions: (i) the evaluator was a member of candidate's thesis committee, (ii) one of them had invited the other to sit in her students' thesis committee, or (iii) both of them sat in the same student thesis committee.

Male candidates tend to have more coauthors among eligible evaluators and they are more likely to have interacted with an eligible evaluator previously in a thesis committee (Table 3, columns 3-5).

3.4 Research similarity

We also collect information on the overlap of research interests between candidates and eligible evaluators. Due to data availability, there are some differences in how we define research similarity in the two countries. In the case of Italy, we have information on the field and the subfield where all researchers with a permanent contract in an Italian university are officially registered. There are 184 fields (*settore concorsuale*) and approximately 370 subfields (*settore scientifico-disciplinare*).¹⁷ Approximately in 60% of the cases the candidate and the eligible evaluator belong to the same subfield (Table 3).

In the case of Spanish researchers, we infer their research interests using information on their participation in doctoral dissertations, either as authors, advisors, or committee members. In Spain, all doctoral theses are classified in more than two thousand categories.¹⁸ Economics, for example, is divided into one hundred different research fields (e.g.: Labor Economics). We construct a measure of the overlap of the research interests of candidates and evaluators based on the subfield of every dissertation where they have been involved. More precisely we construct the following measure:

$$Overlap_{ij} = \sum_c \sqrt{S_i^c S_j^c} \quad (1)$$

where S^c is the ratio of the number of dissertations in category c over the total number of dissertations, in which a certain individual has been involved. This index takes value one if two individuals have participated in dissertations in the same subfields in the

¹⁷Historically, each Italian researcher was assigned to certain *settore scientifico-disciplinare*. More recently, upon the introduction of the new system of competitive exams (*abilitazione scientifica nazionale*) researchers were assigned also to a *settore concorsuale*. The correspondence between the two classifications is not always unique, in some cases researchers belonging to the same *settore scientifico-disciplinare* may be assigned to different *settore concorsuale*.

¹⁸The author of the dissertation selects the subfield using the *International Standard Nomenclature for Fields of Science and Technology*, a system developed by Unesco.

same proportion and value zero if there is no overlap. On average, in our sample the degree of overlap between candidates and evaluators is equal to 0.20. Female candidates are slightly more likely than male candidates to share their research interests with eligible evaluators (Table 3).

4 Empirical analysis

We examine how the gender composition of committees affects the quantity and the quality of male and female candidates who qualify. To achieve a better understanding of the observed patterns, we then explore the potential mechanisms. We examine the role of gender segregation across research networks and across subfields of research, gender stereotypes and taste discrimination. Finally, we use the information provided by individual voting reports in order to explore the interactions that may arise between male and female evaluators

4.1 The impact of committees' gender composition on the chances of success of male and female candidates

We compare the assessments received by male and female applicants and examine how their performance varies with the gender composition of committees. First, we follow an empirical strategy based on observables. Then, we re-examine the data exploiting the random assignment of evaluators to committees.

4.1.1 Descriptive evidence

We estimate the following equation separately for the applicants in the two countries using the ordinary least squares (OLS) method:¹⁹

$$Y_{ie} = \beta_0 + \beta_1 Female_i + \beta_2 Female_i * Female_e + \mathbf{X}_i \beta_3 + \mu_e + \epsilon_{ie} \quad (2)$$

¹⁹Results from probit estimations are very similar and are available upon request. We report the results for the linear probability model because interpreting the interaction effects is simpler.

where Y_{ie} is a dummy variable that takes value one if candidate i qualifies in evaluation e and value zero if the candidate fails to qualify. $Female_i$ is a dummy variable indicating the gender of the candidate and $Female_e$ represents the proportion of women in committee e . $\mathbf{X}_i$ includes observable productivity indicators and individual characteristics. We allow the effect of productivity indicators to vary across disciplinary groups, and the effect of age and contract type to vary across disciplinary groups and levels of promotion. Exam fixed effects (μ_e) control for any differences across exams that might affect the success rate of male and female candidates in a similar way. We cluster standard errors at the committee level.

In Italy, female candidates' success rate is 2.8 percentage points lower than male candidates in the same exam, unconditional on any measure of quality (Table 4, column 1). In Spain the unconditional gender gap is 2.2 percentage points (column 4). Approximately half of the gender gap can be explained by the differences in observable characteristics (columns 2 and 5). It is unclear whether the remaining gap should be attributed to differences in unobservable characteristics or to evaluators discriminatory behavior. Furthermore, the observable individual proxies of quality that we use in our analysis, such as position, affiliation or publications might also be the outcome of discriminatory processes, which would further hinder the interpretation of β_1 .

The gender gap does not decrease when candidates are evaluated by committees including more female evaluators (columns 3 and 6). Actually, in both countries female candidates achieve worse results in committees with more female members, and in Italy this effect is statistically different from zero. Again, these estimates are only indicative and they do not necessarily have a causal interpretation. The gender composition of committees tends to reflect the degree of feminization of the field. There may be substantial differences in the (unobservable) quality of male and female candidates across different fields which are not fully captured by our controls. This may bias our analysis in either direction.

4.1.2 Causal evidence

In order to obtain causal estimates of the impact of committees' gender composition, we exploit the exogenous variation in committee composition provided by the random assignment of evaluators to committees. More precisely, we compare the success rate of male and female candidates who initially were expected to face an evaluation committee with the same gender composition but, due to the random draw, were assigned to committees with different gender compositions. To avoid any potential selection biases, we consider the initial pool of applicants and the initial set of evaluators, independently of whether they eventually withdrew their application or they resigned from the committee. We estimate the following equation using OLS:²⁰

$$Y_{ie} = \beta_0 + \beta_1 Female_i + \beta_2 Female_i * Female_e^{initial} + \beta_3 Female_i * Female_e^{expected} + \mu_e + \epsilon_{ie} \quad (3)$$

where $Female_e^{initial}$ represents the share of female evaluators in the committee that was initially randomly drawn, before any evaluator resigned. $Female_e^{expected}$ is the expected share of women in this committee, calculated based on the composition of the pool of eligible evaluators and the rules that determine the draw. Coefficient β_2 captures the causal effect of committees' gender composition upon the success rate of female candidates, relative to male candidates. Since $Female_e^{initial}$ is computed using the initial assignment of evaluators, coefficient β_2 provides an intention-to-treat estimate. In order to increase the accuracy of the estimation, in some specifications we also include information about individual observable productivity and individual characteristics ($\mathbf{X}_i$). Standard errors are clustered at the committee level.

The causal interpretation of β_2 relies on the assumption that the assignment was indeed random. The way in which the randomization was conducted in each country suggests that there was little room for manipulation.²¹ Nonetheless, before moving into

²⁰Idem.

²¹In Italy, a random sequence of numbers was drawn and was then applied to several disciplines. In Spain, the random draw was carried out publicly on the same day for all disciplines and was certified by the notary.

the discussion of the impact of committees' gender composition on candidates' chances of success, we verify empirically that, conditional on the expected composition of the committee, its actual composition is uncorrelated with any observable predetermined factor. We estimate equation (3) using the eleven predetermined variables that are common for Italian and Spanish databases. As expected, the evidence is consistent with the assignment being indeed random. Out of twenty two coefficients, only one is significantly different from zero at 5% level (Table 5). A joint F-test cannot reject that the difference in quality between female and male candidates is similar across committees with different gender compositions.

Next, we examine the causal impact of committees' gender composition. The estimates are in line with our preliminary results based on observables. In Italy, the proportion of women in committees has a significant negative impact on the relative chances of success of female candidates (Table 6, column 1). In Spain, the effect of female evaluators on the relative success of female candidates is also negative, though it is not significantly different from zero (column 5). These estimates are (statistically) unchanged when we include in the estimation the available information on candidates' research output and other observable characteristics (columns 2 and 6). In columns 3 and 7, we take into account that some evaluators declined to participate in committees. We instrument the final gender composition of the committee using the initial composition determined by the random draw. The results are very similar; if anything, the impact of committees' gender composition is slightly greater. In quantitative terms, in Italy an additional female evaluator decreases the relative chances of success of female candidates by approximately 2.6 percentage points ($\Delta Female_e=1/5$). In the Spanish committees, the effect is around 0.2 p.p. ($\Delta Female_e=1/7$; $\beta_2 = 0.016$).

To make these two estimates more comparable, it is useful to express them taking into account the average success rate of female candidates in each country and to consider explicitly the upper and the lower bounds of a 95% confidence interval. In Italy, an extra woman on the committee lowers the success rate of women by somewhere between 4% and 11%. In Spain, an extra woman on the committee can lower the success

rate of women by at maximum 9%, but she can also increase it by up to 5%. In sum, the impact that women in committees have upon the success rate of female candidates is more negative and more precisely estimated in the Italian case, but we cannot reject that the effect is statistically similar in the two countries.

The range of variation in gender composition that we exploit in our analysis is typically between committees with no women and committees with a minority of women. In Appendix B we also show that within this range there are no significant non-linearities.

4.2 Does the presence of women in the committee affect candidates' decision to apply?

So far we have considered the sample of all initial candidates. Some of these candidates dropped from the evaluation process after committees were formed, perhaps because they anticipated that they had a small chance to qualify and they preferred to avoid the costs associated to failure, and they did not receive an evaluation from the committee.

Therefore, the above estimates may in principle capture the effect that the gender composition of a committee has upon candidates' decision to self-select into the process. To examine this issue, we use data from Italy and estimate equation (2) using as the dependent variable the indicator for those potential candidates who did not withdraw their application. While relatively fewer women decided to go ahead with the application (-2.6 p.p.), these differences are not related to the share of female evaluators (Table 6, column 4). The evidence thus suggests that impact of committees' gender composition on the chances of success of candidates can only be attributed to evaluations.

4.3 Does the presence of women in the committee affect the quality of promoted candidates?

An additional justification for increasing female representation in committees might be that female researchers help to reduce evaluation biases and select better candidates,

even though not necessarily more female candidates. To shed light on this issue, we compare the observable quality of candidates who qualified in committees with different gender compositions:

$$x_{ie} = \beta_0 + \beta_1 Female_e + \beta_2 Female_e^{expected} + \epsilon_{ie} \quad (4)$$

where x_{ie} is a proxy of candidate i 's quality, measured at the time of the evaluation or during the following five years. We instrument the final gender composition of the committee ($Female$) using the original one ($Female^{initial}$), and we cluster standard errors at the committee level.

We consider several proxies of quality. First, we consider the research output of successful candidates at the time of the evaluation. As shown in Table 7, candidates that were promoted by committees with a different gender composition are at the time of the evaluation statistically similar in terms of the number of papers that they have published, the quality of the journals, the number of students advised or their participation in theses committees.

Using the Spanish data, we also examine the research productivity of successful candidates during the five-year period following the evaluation. Additionally, for the candidates who qualified to positions of associate professor, we check whether they succeeded in obtaining a qualification for full professorship. Once again, we see no evidence that the quality of candidates who qualify is related to the number of women who sat on these candidates' evaluation committees. Overall, we do not observe any indication that committees with more female evaluators select better or worse candidates.

4.4 Mechanisms

The two large-scale randomized natural experiments provide a clear result: increasing the proportion of women in scientific committees does not increase the average success rate of female candidates relative to the success of male candidates. Below, we try to provide a more in-depth examination of the main theoretical arguments discussed

in the literature supporting a larger presence of women in committees, and we try to understand why they are not empirically relevant.

4.4.1 Gender segregation across research networks

One of the arguments behind gender quotas is the existence of ‘old-boy networks’. If professional connections with committee members help to achieve success and, at the same time, these connections are gendered, female candidates might be at a disadvantage when evaluation committees do not include women.

Previous studies suggest that the presence of strong connections in evaluation committees, as measured by co-authorships, researchers’ affiliation, PhD supervisions and participation in doctoral theses committees, have a positive impact on candidates’ chances of success (Zinovyeva and Bagues 2015, Bagues, Sylos-Labini and Zinovyeva 2015). The empirical evidence of the previous section shows that the presence of more women in the committee does not increase the relative success rate of female candidates. Is this because in Spain and Italy research networks are not gendered?

We consider all possible pairs between candidates and potential evaluators within a given field and we analyze whether the probability of being linked varies with their gender:

$$Link_{ij} = \beta_0 + \beta_1 Female_i + \beta_2 Female_j + \beta_3 Female_i * Female_j + \mu_e \beta_4 + \epsilon_{ij} \quad (5)$$

where $Link_{ij}$ stands for any of the observable links between candidate i and potential evaluator j . $Female_i$ and $Female_j$ are indicators for female candidates and evaluators, and μ_e are exam fixed effects. Coefficient β_3 in this equation reflects whether female candidates are more likely to be connected with female eligible evaluators than with male ones.

Links are gendered in every observable dimension (Table 8). There is gender segregation across institutions. In Italy, the likelihood of observing a female candidate with the same affiliation as a female professor is 0.3 p.p. (13%) larger than the likelihood of observing a similar link between a female candidate and a male professor. In

Spain, female candidates are 0.4 p.p (9%) more likely to be in the same institution as a female professor, relative to the probability of being affiliated to the same institution as a male professor. Co-authorships are also more likely when individuals share the same sex. In Italy female candidates are 0.2 p.p (19%) more likely to co-author with a female eligible evaluator than with a male one; in Spain the premium is equal to 0.1 p.p. (22%). Similarly, PhD supervisions and participation in PhD committees are also gendered. Female candidates are 0.2 p.p. (20%) more likely to have a female advisor and 1.4 p.p. (10%) more likely to have participated in the same dissertation committee as a female eligible evaluator.

We have shown that there is gender segregation across research networks. Next, we study whether candidates benefit from the presence of a connected evaluator in the committee. Specifically, we estimate the following equation:

$$Y_{ie} = \beta_0 + \beta_1 Female_i + \beta_2 Female_i * Female_e + Links_{ie} \beta_3 + \beta_5 Female_i * Female_e^{expected} + Links_{ie}^{expected} \beta_6 + \mathbf{X}_i \beta_8 + \mu_e + \epsilon_{ie} \quad (6)$$

where $Links_{ie}$ is a vector including the different types of links between committee members and candidates. We also include as controls the expected proportion of links in the committee and we instrument the final composition of the committee ($Female_e$, $Links_{ie}$) using the outcome of the initial lottery draw. The vector of coefficients β_3 provides information about the causal impact of connections in the committee.

Table 9 reports the results of this analysis. Connections with evaluators are important for promotion. The presence of a colleague in the committee increases the success rate of connected candidates by 3.6 p.p. (10%) in Italy and by 4.5 p.p. (41%) in Spain. The impact of co-authors is slightly larger: 5.0 p.p. (14%) in Italy and 12.5 p.p. (113%) in Spain. Candidates with a advisor in the evaluation committee also enjoy a premium of 9.0 p.p. (82%) and when an evaluator has interacted previously with the candidate in some thesis committee the premium is around 2.5 p.p. (22%). These results are similar to the ones reported in Zinovyeva and Bagues (2015) and Bagues, Sylos-Labini and Zinovyeva (2015).

While connections are gendered and their impact is large, their inclusion as controls in the analysis does not affect significantly our estimates of the effect of evaluators' gender on candidates' success rate (columns 1 and 5 vs. columns 2 and 6). A plausible explanation for why connections, while being gendered, do not affect significantly our estimates may be related to their scarcity in a context of evaluations at the national level. For instance, in Italy the probability that a female candidate and a male evaluator are co-authors is around 1.4%. This probability increases by 0.2 p.p. when the evaluator is also female, which would translate into an increase in the success rate of female candidates by a mere 0.01 p.p.²² Moreover, as we show in Appendix C, evaluators' support of connected candidates does not depend on their gender.

It is also possible that weaker links between candidates and evaluators, not considered above, are also gendered (for instance, co-authors of co-authors). Nonetheless, the analysis performed by Zinovyeva and Bagues (2015) suggests that these indirect links do not affect evaluation outcomes.

4.4.2 Gender segregation across research subfields

If committee members tend to prefer candidates with similar research interests and, at the same time, men and women are segregated across research subfields, the lack of women in committees might hinder the ability of female candidates to succeed.

We check whether candidates are more likely to have the same research interests as eligible evaluators of the same gender. We estimate equation (5) using as the dependent variable the research similarity between candidates and eligible evaluators. We find gender segregation across research subfields in both countries but its magnitude is relatively small. In Italy, female candidates are 1.7 p.p. (3.5%) more likely to be in the same subfield as a female professor. In Spain, the overlap between female candidates and female eligible evaluators is 0.4 p.p. (2%) larger (Table 8, columns 3 and 8).

Research similarity with evaluators tends to increase candidates' chances of suc-

²²A back of the envelope calculation suggests that a 0.2 p.p. increase in the probability of having a coauthor in the committee, times the premium associated to the presence of a coauthor in the committee (5.0 p.p.), is equal to 0.01 p.p.

cess, but the effect of female evaluators on female candidates' relative success rate is unchanged when we control in the estimation for research similarity (Table 9). This is consistent with the relatively small level of gender segregation observed within fields. In sum, the gender segregation across research interests is too limited for female candidates to benefit significantly from more female evaluators in the committee.

4.4.3 Stereotypes

An additional theoretical argument in favor of a higher female presence in evaluation committees is that senior male researchers might have stereotypes against female candidates. If senior female researchers do not share these stereotypes, having more women on the committee might reduce the impact of gender prejudices.

Stereotyping might be stronger when evaluators are less informed about candidates' quality. It might be particularly difficult to assess the quality of candidates who do research in subfields that lie far away from evaluators' knowledge. To investigate this issue, we divide evaluations in two groups based on the distance between evaluators' and candidates' research interests. When candidates and evaluators work in similar areas, evaluators' gender is irrelevant (Table 10, first row). However, when candidates do research in a different subfield, female candidates tend to perform relatively worse when there are relatively more women in the committee. This pattern is observed in both countries. If anything, the evidence is more consistent with stereotypes against women being more relevant in gender-mixed committees than in all-male committees.

It is also sometimes argued that stereotyping against women is stronger in sciences and mathematics-related disciplines (Reuben, Sapienza, and Zingales 2014). We compare the effect of female evaluators in STEMM and SSH disciplines, but we do not observe any significant differences between these two groups neither in Spain nor in Italy (second row of Table 10).

One might also expect prejudices against women to be stronger in disciplines that are less feminized and, therefore, offer fewer chances to interact with female researchers. We examine separately disciplines with a relatively low and a relatively high proportion

of female full professors. We do not find any evidence suggesting that evaluators in these two groups differ in terms of their preference for candidates of the same sex (third row of Table 10).

4.4.4 High-level positions

The impact of committees' gender composition might also depend on the importance of the position at stake. Taste-based discrimination against women might be stronger when female candidates aspire to a high-level position. Some male evaluators might be reluctant to see a female colleague at the top of the academic career ladder. They might also hold negative stereotypes of women, for instance, regarding their leadership or other abilities specific to full professor positions.

We examine separately the effect of female presence upon the evaluation committee for candidates to full and associate professor positions (fourth row of Table 10). We do not observe any significant differences between these groups of evaluations in Italy, but we do observe a significant difference between exams for full and associate professorships in Spain. Specifically, it appears that in Spain, in committees assessing candidates to full professor positions, a higher female presence has a positive impact on female candidates' relative chances of success. However, the opposite is true in evaluations for promotion to more junior positions.

So, in the case of promotions to full professorships in Spain, but not in Italy, the result is consistent with the existence of stereotypes, or even of taste discrimination, against women by committees with low or no representation of women.

4.4.5 Analysis by disciplinary groups

Beyond these theories, it might be that the gender composition of committees matters in some specific fields. The previous empirical literature on the impact of evaluators' gender does not provide a clear pattern. Two articles conducted in Science and Economics find that evaluators tend to prefer candidates of the same sex (Casadevall and Handelsman 2013 and De Paola and Scoppa 2015) but in two other studies conducted

in the same disciplines evaluators exhibit a preference for candidates of the other sex (Broder 1993, Ellemers et al. 2004). Five other articles in different fields do not find any significant relationship, but estimates are in general not precise.²³

We consider 16 different groups of disciplines: Civil Engineering, Architecture, Geology, Social Sciences, Psychology, Veterinary, Physics, Chemistry, Mathematics, History, Medicine, Biology, Economics and Business, Law, Languages and Industrial Engineering. We estimate equation (3) separately for each group and each country. We report these estimates in Figure 1. Out of 32 coefficients, 21 are not significant, 5 are significantly positive and 6 are significantly negative. However, when we take into account in the calculation of standard errors that we are running multiple regressions using a Bonferroni correction none of the coefficients remains significant. Altogether, it is not possible to reject that the impact is similar to zero in any of the different fields and countries.

Figure 1 also illustrates that estimations on small samples tend to produce estimates of an excessive magnitude (Gelman and Weakliem 2009). In the figure groups are ordered according to their size, from smaller to larger. Estimates tend to be less precise and also larger in absolute terms in the left-hand side of the figure. As the number of available observations in the field increases, the estimate becomes more accurate and also smaller in absolute terms.

4.5 Committee decision-making

So far we have documented that mixed-gender committees are not more favorable towards female candidates than all-male committees. There are several possible explanations. It might be that female evaluators are as biased (or unbiased) as male evaluators. Alternatively, maybe female evaluators are more favorable towards female candidates but they are in a minority and they are unable to have any impact on the decision of the committee. Their presence in the committee might also induce male evaluators to be less favorable towards female candidates.

²³See more details in Table A1.

To shed light on this issue, we compare the individual evaluations casted by male and female evaluators using the information provided by around 300,000 individual voting reports. We estimate the following equation:

$$V_{ij} = \beta_0 + \beta_1 Female_j + \beta_2 Female_i * Female_j + \mu_i + \epsilon_{ij}, \quad (7)$$

where V_{ij} is a variable that takes value one if evaluator j casted a positive vote for candidate i , while $Female_i$ and $Female_j$ are indicators that capture the gender of the candidate and the evaluator respectively. A vector of candidates' fixed effects μ_i captures any differences in candidates' characteristics that are observable to all evaluators. Female and male evaluators exhibit the same grading standards with male candidates. Female evaluators are 0.6 p.p. (1.3%) more likely to vote in favor of female candidates than male evaluators, but this difference is not statistically different from zero (Table 11, column 1).

Another question that we would like to answer is whether the voting behavior of male evaluators changes when there are more female evaluators on committee. We estimate the following equation to explicitly address this question:

$$\begin{aligned} V_{ije} = & \beta_0 + \beta_1 Female_i + \beta_2 Female_j + \beta_3 Female_i * Female_j + \\ & + [\gamma_1 Female_i + \gamma_2 Female_j + \gamma_3 Female_i * Female_j] * Female_e^{final} + \\ & + \mathbf{X}_{ie} \beta_4 + \epsilon_{ije}, \end{aligned} \quad (8)$$

where $Female_e^{final}$ represents the share of female evaluators in the evaluation committee. Coefficient β_1 captures whether, in all-male committees, female candidates receive fewer positive votes than male candidates of comparable quality and coefficient γ_1 shows whether this gender gap in votes casted by male evaluators changes when there are female evaluators in the committee. According to our estimates, in all-male committees, male and female evaluators receive similar evaluations ($\beta_1 = 0.004$, st. error=0.007). Each additional female evaluator in the committee decreases the probability that a female candidate receives a positive vote from a male evaluator by 1 p.p.,

an effect which is marginally significant ($\Delta = 1/5$, $\gamma_1 = -0.048$, st. error=0.028).

5 Conclusions

A larger presence of women in committees is frequently defended in policy discussions. In this paper we analyze how the gender composition of scientific committees affects the chances of success of female and male candidates using the exceptional evidence provided by qualification evaluations for full and associate professorships in every discipline in two different countries, Italy and Spain. These evaluations involved around 100,000 applications and 8,000 evaluators. The random assignment of evaluators to committees creates a setting of large-scale natural randomized experiments.

In general, the presence of female evaluators in the committee neither increases the success rate of female candidates nor it improves the quality of selected candidates. Strikingly, in most subsamples we observe the opposite pattern: committees with more women tend to be less favorable towards female candidates. The only exception are evaluations to full professorships in Spain, where female candidates have better chances of success when they are evaluated by a committee with more women.

We explore why the standard theoretical arguments that are usually employed in support of a higher representation of women in scientific committees do not hold in this context. We observe some gender segregation in research interests, but its magnitude is relatively small. Women are 2-4% more likely to do research in the same subfield as a female evaluator than a male evaluator. Evaluators tend to have a preference for candidates with similar research interests but overall the impact of gender segregation across research subfields on evaluation outcomes is very limited.

We also document that in Italy and in Spain there exists gender segregation across research networks. Female candidates are significantly more likely to be connected to female evaluators than to male evaluators as measured by co-authorships, affiliation, thesis advisor and participation in theses committees. This difference is large (10%-20%) but, in the context of the nation-wide evaluations that we consider in this paper, the likelihood of having a connection in the evaluation committee is small and thus the

impact of gendered networks on evaluations is again limited.

Another justification for increasing the presence of women in committees is that male evaluators may hold stereotypes that have a negative effect upon female candidates. In order to explore the potential impact of gender stereotypes, we focus on cases where evaluators are not familiar with the research profile of the candidate. When information asymmetries are important, gender-mixed committees are less favorable towards women than all-male committees. One possible explanation is that female evaluators hold stronger stereotypes against women than male evaluators. Another interpretation of the evidence is that the presence of women in the committee affects male evaluators behavior. As soon as a woman colleague joins the committee, men perhaps reduce their commitment for gender equality, or it might also be that their male identity is strengthened (Akerlof and Kranton 2000). Information from individual votes within committees provides support for the later hypothesis. Female evaluators are, if anything, relatively more favorable towards female candidates, and male evaluators become less favorable towards female candidates when women are present in the committee, although this result is only significant at the 10% level.

Several countries have introduced quotas in scientific committees requiring the presence of a minimum share of male and female evaluators. There are certain features of quotas that are not captured by our analysis. Evaluators that are explicitly chosen to represent a minority might behave differently, perhaps being more inclined to take a positive view of candidates belonging to their own group. Moreover, the introduction of quotas may affect the strategic incentives of evaluators. Nonetheless, keeping in mind these limitations, our results cast doubts on a generalized implementation of gender quotas in scientific committees. According to our results, a higher representation of women in scientific committees neither increases the number of promoted female candidates nor helps to promote candidates that prove to be more productive in the future, and it might also have unintended consequences. Quotas may be detrimental for senior female researchers, who would have to spend more time sitting on committees and, in some cases, for junior ones, who may even experience a reduction in their chances of

success.

References

- Abe, Yukiko (2012), "The Academic Labor Market in Japan and the Presence of Women," *CSWEP Newsletter*, Fall, pp. 9-10.
- Abrevaya, Jason and Daniel S. Hamermesh (2012), "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?," *Review of Economic and Statistics*, Vol. 94(1), pp. 202-207.
- Akerlof, George A. and Rachel E. Kranton (2000), "Economics And Identity," *The Quarterly Journal of Economics*, MIT Press, Vol. 115(3), pp. 715-753.
- Babcock, Linda, Michele Joy Gelfand, Deborah Small, and Heidi Stayn (2006), "Propensity to initiate Negotiations: A New look at gender variation in Negotiation Behavior, in David De Cremer, Marcel Zeelenberg, and J. Keith Murnighan (Eds.) *Social Psychology and Economics*. Mahwah, NY: Lawrence Erlbaum Associates, pp. 239-259.
- Bagues, Manuel and Berta Esteve-Volart (2010), "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *Review of Economic Studies*, Vol. 77(4), pp. 1301-1328.
- Bagues, Manuel and Maria Jose Perez-Villadoniga (2012), "Do recruiters prefer applicants with similar skills? Evidence from a randomized natural experiment," *Journal of Economic Behavior and Organization*, Vol. 82(1), pp. 12-20.
- Bagues, Manuel and Maria Jose Perez-Villadoniga (2013), "Why do I like people like me?," *Journal of Economic Theory*, Vol. 148(3), pp. 1292-1299.
- Bagues, Manuel, Mauro Sylos-Labini and Natalia Zinovyeva (2015), "The Hidden Advantage of Connections in Scientific Committees: Evidence from a Large Scale Randomized Natural Experiment," Aalto University, mimeo.
- Barres, Ben (2006), "Does gender matter?" *Nature*, Vol. 442, pp. 133-136.
- Blackaby, David, Alison L. Booth and Jeff Frank (2005), "Outside Offers And The Gender Pay Gap: Empirical Evidence From the UK Academic Labour Market," *Economic Journal*, Vol. 115(501), pp. F81-F107.
- Blau, Francine D., Janet M. Curie, Rachel T.A. Croson and Donna K. Ginther (2010), "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial," *American Economic Review*, Vol. 100(2), pp. 348-352.
- Booth, Alison L. and Andrew Leigh (2010), "Do Employers Discriminate by Gender? A Field Experiment in Female-Dominated Occupations," *Economic Letters*, Vol. 107(2), pp. 236-238.
- Boschini, Anne and Anna Sjögren (2007), "Is Team Formation Gender Neutral? Evidence from Coauthorship Patterns," *Journal of Labor Economics*, Vol. 25(2), pp. 325-365.
- Bosquet, Clément, Pierre-Philippe Combes, Cecilia Garcia-Peñalosa (2013), "Gender and Competition: Evidence from Academic Promotions in France," *CESifo Working Paper*, No. 4507.

- Brescoll, Victoria L. (2011), "Who Takes the Floor and Why Gender, Power, and Volubility in Organizations, *Administrative Science Quarterly*, Vol. 56(4), pp. 622-641.
- Broder, Ivy E. (1993), "Review of NSF Economics Proposals: Gender and Institutional Patterns," *American Economic Review*, Vol. 83(4), pp. 964-970.
- Buser, Thomas, Muriel Niederle and Hessel Oosterbeek (2014), "Do Women Shy away from Competition? Do Men Compete too Much?" *Quarterly Journal of Economics*, Vol. 129(3), pp. 1409-1447.
- Casadevall, Arturo and Jo Handelsman (2014), "The Presence of Female Conveners Correlates with a Higher Proportion of Female Speakers at Scientific Symposia," *mBio*, Vol. 5(1), e00846-13.
- Ceci, Stephen J. and Wendy M. Williams (2011), "Understanding current causes of women's underrepresentation in science," *Proceedings of the National Academy of Sciences*, Vol. 108(8), pp. 3157-3162.
- De Paola, Maria and Vincenzo Scoppa (2015), "Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy," *Economica*, Vol. 82 (325), pp. 162-188.
- De Paola, Maria, Michela Ponzo and Vincenzo Scoppa (2015), "Gender Differences in Attitudes Towards Competition: Evidence from the Italian Scientific Qualification," *IZA discussion paper*, No. 8859.
- Dolado, Juan Jose, Florentino Felgueroso and Miguel Almunia (2012), "Are men and women-economists evenly distributed across research fields? Some new empirical evidence," *SERIEs: Journal of the Spanish Economic Association*, Vol. 3(3), pp. 367-393.
- Ellemers, Naomi, Henriette Van den Heuvel, Dick de Gilder, Anne Maass, and Alessandra Bonvini (2004), "The underrepresentation of women in science: Differential commitment or the queen bee syndrome?" *British Journal of Social Psychology*, Vol. 43 (3), pp. 315-338.
- European Commission (2008), *Mapping the Maze: Getting More Women to the Top in Research*, Luxembourg: Publication Office of the European Union.
- European Commission (2013), *She Figures 2012: Gender in Research and Innovation*, Luxembourg: Publication Office of the European Union.
- Fuentes, Eulàlia and Llorenç Arguimbau (2010), "Las tesis doctorales en España (1997-2008): análisis, estadísticas y repositorios cooperativos," *Revista Española de Documentación Científica*, Vol. 33(1), pp. 63-89.
- Gelman, Andrew and David Weakliem (2009), "Of beauty, sex, and power: Too little attention has been paid to the statistical challenges in estimating small effects," *American Scientist*, Vol. 97 (4), pp. 310-316.
- Giménez-Toledo, Elea, Carlos Tejada-Artigas and Jorge Mañana-Rodríguez (2012), "Scholarly Publishers Indicators (SPI), 1st edition. Available at: <http://epuc.cchs.csic.es/SPI> [Accessed Nov. 18 2013.]
- Ginther, Donna K. and Shulamit Kahn (2004), "Women in Economics: Moving Up or Falling Off the Academic Ladder," *Journal of Economic Perspectives*, Vol. 18(3), pp 193-214.

- Ginther, Donna K. and Shulamit Kahn (2009), “Does Science Promote Women? Evidence from Academia 1973-2001,” NBER chapters in: *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, pp. 163-194.
- Hale, Galina and Tali Regev (2014), “Gender Ratios at Top PhD Programs in Economics,” *Economics of Education Review*, Vol. 41, pp. 55-70.
- Hilmer, Christiana and Michael Hilmer (2007), “Women Helping Women, Men Helping Women? Same-Gender Mentoring, Initial Job Placements, and Early Career Publishing Success for Economics PhDs,” *American Economic Review*, Vol. 97(2), pp. 422-426.
- Jayasinghe, Upali W., Herbert W. Marsh and Nigel Bond (2003), “A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 166 (3), pp. 279-300.
- Karpowitz, Christopher F., Tali Mendelberg and Lee Shaker (2012), “Gender Inequality in Deliberative Participation,” *American Political Science Review* Vol. 106(3), pp. 533-547.
- National Research Council (2009), *Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty*, Washington D.C.: The National Academies Press.
- Maraut, Stéphane and Catalina Martínez (2014), “Identifying author-inventors from Spain: methods and a first insight into results,” *Scientometrics*, Vol. 101, pp. 445-476.
- Méndez, María J and John R. Busenbark (2015), “Shared leadership and gender: all members are equal... but some more than others”, *Leadership and Organization Development Journal*, Vol. 36(1), pp. 17-34.
- Niederle, Muriel and Lise Vesterlund (2007), “Do women shy away from competition? Do men compete too much?” *Quarterly Journal of Economics*, Vol. 122 (3), pp. 1067-1101.
- Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham and Jo Handelsman (2012), “Science faculty’s subtle gender biases favor male students,” *Proceedings of the National Academy of Sciences*, Vol. 109(41), pp. 6474-16479.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2014), “How stereotypes impair women’s careers in science,” *Proceedings of the National Academy of Sciences*, Vol. 111 (12) 4403-4408, doi: 10.1073/pnas.1314788111.
- Sandberg, Anna (2014), “Competing biases: Effects of gender and nationality in sports judging, Stockholm School of Economics, mimeo.
- Smith, Kristin A., Paola Arlotta, Fiona M. Watt, The Initiative on Women in Science and Engineering Working Group, and Susan L. Solomon (2015), “Seven Actionable Strategies for Advancing Women in Science, Engineering, and Medicine,” *Cell Stem Cell*, Volume 16(3), pp. 221-224.
- Staines, Graham, Carol Tavis and Toby Jayaratne (1974), “The queen bee syndrome,” *Psychology Today*, 7(8), pp. 63-66.

- Steinpreis, Rhea E. and Katie A. Anders, and Dawn Ritzke (1999), "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study," *Sex Roles*, Vol. 41(7/8), pp. 509-528.
- Torres-Salinas, Daniel, Maria Bordons, Elea Giménez-Toledo, Emilio Delgado-López-Cózar, Evaristo Jiménez-Contreras y Elías Sanz-Casado (2010), "Clasificación integrada de revistas científicas (CIRC): propuesta de categorización de las revistas en ciencias sociales y humanas", *El profesional de la información*, v. 19, n. 6, pp. 675-683.
- van den Brink, Marieke, Yvonne Benschop and Willy Jansen (2010), "Transparency in Academic Recruitment: A Problematic Tool for Gender Equality?" *Organization Studies*, Vol. 31 (11), pp. 1459-1483.
- Vesterlund, Lise, Linda Babcock, and Laurie Weingart (2014), "Breaking the Glass Ceiling with "No": Gender Differences in Declining Requests for Non-Promotable Tasks", mimeo.
- Vernos, Isabelle (2013), "Research management: Quotas are questionable, *Nature*, Vol. 495(7439), pp. 39.
- Williams, Wendy M. and Stephen J. Ceci (2015), "National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track," *Proceedings of the National Academy of Sciences*, Vol 112 (17), pp. 5360-5365.
- Woolley, Anita Williams, Chabris, Christopher F., Pentland, Alex, Hashmi, Nada, and Thomas W. Malone (2010), "Evidence for a Collective Intelligence Factor in the Performance of Human Groups," *Science*, Vol. 330, pp. 686-688.
- Zinovyeva, Natalia and Manuel Bagues (2015), "The role of connections in academic promotion," *American Economic Journal: Applied Economics*, Vol. 7(2), pp. 264-292.

Table 1: Descriptive statistics – Eligible evaluators

	1	2	3	4	5	6	7	8	9	10	11	12
	Italy				Spain							
	Full professors				Full professors				Associate professors			
	All	Male	Female	p-value	All	Male	Female	p-value	All	Male	Female	p-value
Female	0.20				0.14				0.35			
Tenure in position	13	0.05	-0.24	0.000	13	0.05	-0.33	0.000	10	0.01	-0.03	0.000
Age	-				52	0.00	-0.03	0.010	45	0.01	-0.02	0.001
All Publications:	131	0.03	-0.14	0.000	34	0.02	-0.14	0.000	14	0.05	-0.09	0.000
- Articles	72	0.04	-0.18	0.000	30	0.02	-0.14	0.000	12	0.05	-0.10	0.000
- Books	5	0.04	-0.16	0.000	1	0.01	-0.06	0.000	0.44	0.01	-0.02	0.000
- Book chapters	20	0.01	-0.06	0.004	3	0.01	-0.07	0.000	1	0.01	-0.01	0.003
- Conference proceedings	23	-0.00	0.01	0.545	-				-			
- Patents	0.42	0.00	-0.01	0.527	0.10	0.00	-0.02	0.001	0.04	0.01	-0.02	0.000
- Other	11	0.02	-0.07	0.001	-				-			
Total Article Influence Score	132	0.03	-0.20	0.000	33	0.01	-0.10	0.000	12	0.03	-0.07	0.000
A-journal articles	12	0.04	-0.12	0.000	4	0.03	-0.15	0.000	2	0.04	-0.06	0.000
PhD students advised	-				5	0.03	-0.20	0.000	1	0.08	-0.15	0.000
PhD committees	-				25	0.05	-0.33	0.000	5	0.07	-0.13	0.000
Observations	5,876				49,199				61,052			

Notes: The table provides descriptive information for the pool of eligible evaluators in qualification exams in Italy and in Spain. In Italy it includes only evaluators who are based in an Italian university. *Article Influence Score* is only available for candidates in science, technology, engineering, mathematics and medicine. Information on publications in *A-journal articles* is only provided for candidates in social sciences and humanities. Columns 1, 5 and 9 report mean values for each corresponding variable and sample. In columns 2, 3, 6, 7, 10 and 11 variables have been normalized to have zero mean and unit variance for individuals within each field and rank. Columns 4, 8 and 12 report the p-value of a t-test of the difference in means between male and female eligible evaluators in the corresponding variable.

Table 2: Descriptive statistics – Applications

	1	2	3	4	5	6	7	8	9	10
	Applications to full professorships					Applications to associate professorships				
	Mean	St.Dev.	Male	Female	p-value	Mean	St.Dev.	Male	Female	p-value
Italy										
Female	0.31	0.46				0.41	0.49			
Age	49	8	-0.01	0.01	0.205	43	7	0.02	-0.03	0.000
Based in a national university	0.96	0.19	0.96	0.98	0.000	0.96	0.18	0.96	0.97	0.000
Permanent position:	0.74	0.44	0.72	0.77	0.000	0.47	0.50	0.46	0.48	0.000
- same field	0.77	0.42	0.76	0.80	0.000	0.74	0.44	0.72	0.76	0.000
Application order	0.50	0.29	0.50	0.51	0.012	0.50	0.29	0.50	0.50	0.717
CV length (pages)	20	79	-0.01	0.03	0.006	14	60	-0.03	0.04	0.000
All Publications:	89	83	0.04	-0.09	0.000	53	54	0.04	-0.06	0.000
- Articles	52	65	0.06	-0.14	0.000	29	41	0.07	-0.10	0.000
- Books	2	4	0.04	-0.09	0.000	1	2	0.05	-0.08	0.000
- Book chapters	9	13	0.01	-0.03	0.000	5	8	0.01	-0.02	0.004
- Conference proceedings	17	32	-0.01	0.03	0.002	11	21	-0.01	0.02	0.000
- Patents	0.35	2.09	0.01	-0.03	0.000	0.19	1.39	0.03	-0.04	0.000
- Other	8	21	0.01	-0.02	0.037	6	16	0.01	-0.01	0.048
Number of coauthors per article	5	6	-0.01	0.03	0.003	5	6	-0.03	0.04	0.000
First-authored	0.22	0.19	-0.01	0.02	0.039	0.22	0.2	0.00	-0.01	0.324
Last-authored	0.15	0.17	0.02	-0.04	0.000	0.11	0.16	0.02	-0.03	0.000
Average Article Influence Score	1.28	0.93	0.03	-0.09	0.000	1.28	0.98	0.03	-0.04	0.000
A-journal articles	6	10	0.04	-0.08	0.000	4	6	0.03	-0.04	0.000
Qualified	0.36	0.48	0.37	0.34	0.000	0.37	0.48	0.38	0.35	0.000
Failure	0.48	0.50	0.48	0.46	0.013	0.50	0.50	0.5	0.5	0.969
Withdrawal	0.16	0.37	0.15	0.20	0.000	0.13	0.34	0.12	0.16	0.000
Proportion of positive votes	0.47	0.46	0.47	0.46	0.242	0.45	0.47	0.46	0.44	0.000
Number of applications	21,594					47,426				
Spain										
Female	0.27	0.44				0.40	0.49			
Age	46	6	-0.01	0.03	0.015	37	6	0.03	-0.05	0.000
All Publications:	19	21	0.03	-0.09	0.000	8	14	0.07	-0.10	0.000
- Articles	17	21	0.04	-0.09	0.000	7	14	0.07	-0.11	0.000
- Books	0.64	1.47	0.01	-0.03	0.005	0.21	0.65	0.02	-0.02	0.000
- Book chapters	1.57	3.18	0.01	-0.02	0.086	0.54	1.41	0.01	-0.01	0.025
- Patents	0.04	0.33	0.00	0.00	0.919	0.02	0.22	0.01	-0.01	0.012
Number of coauthors per article	3	10	0.00	0.01	0.691	5	23	0.00	0.00	0.863
First-authored	0.25	0.31	0.00	0.00	0.862	0.26	0.34	0.01	-0.01	0.200
Last-authored	0.24	0.30	0.01	-0.02	0.220	0.17	0.30	0.03	-0.05	0.000
Average Article Influence Score	0.75	0.43	-0.01	0.02	0.458	0.72	0.54	0.03	-0.06	0.000
A-journal articles	3	5	0.05	-0.10	0.000	1	2	0.06	-0.06	0.000
PhD students advised	2	3	0.03	-0.09	0.000	0.24	0.88	0.03	-0.05	0.000
PhD committees	7	9	0.03	-0.08	0.000	1	3	0.05	-0.08	0.000
Qualified	0.11	0.31	0.11	0.09	0.003	0.12	0.32	0.12	0.11	0.025
Number of applications	13,444					17,799				

Notes: *Article Influence Score* is only available for candidates in science, technology, engineering, mathematics and medicine. Information on publications in *A-journal articles* is only provided for candidates in social sciences and humanities. Columns 1 and 6 report mean values for each corresponding variable and sample. In columns 3, 4, 8, and 9 all productivity variables have been normalized to have zero mean and unit variance for applications within each exam. Columns 5 and 10 report the p-value of a t-test of the difference in means between male and female candidates in the corresponding variable.

Table 3: Descriptive statistics – Links and Research Overlap

	1	2	3	4	5
	All		Male	Female	
<i>Italy</i>	N	Mean	Mean	Mean	p-value
Colleagues	2,555,839	0.028	0.027	0.030	0.000
Coauthors	2,555,839	0.014	0.015	0.014	0.000
Same subfield	1,373,790	0.598	0.597	0.599	0.022
<i>Spain</i>					
Colleagues	5,445,067	0.046	0.047	0.043	0.000
Coauthors	5,445,067	0.004	0.005	0.004	0.000
PhD advisor	5,445,067	0.002	0.002	0.002	0.322
PhD thesis committee	5,445,067	0.013	0.014	0.011	0.000
Overlap in research interests	4,711,621	0.201	0.189	0.221	0.000

Notes: The table provides information on links between candidates and eligible evaluators within each discipline. Information about research interests is only available for candidates with a permanent contract in an Italian university and for candidates who have defended their thesis in Spain or who have participated in a thesis committee in Spain. The variable *Same subfield* takes value one if a candidate and an eligible evaluator belong to the same subfield (*settore scientifico-disciplinare*). The variable *Overlap in research interests* measures the degree of overlap between the research interests of eligible evaluators and candidates, as measured by their participation in PhD thesis committees.

Table 4: Success rate by gender of candidates and evaluators

	1	2	3	4	5	6
	Italy			Spain		
Female	-0.028*** (0.006)	-0.016*** (0.005)	-0.006 (0.007)	-0.022*** (0.004)	-0.014*** (0.004)	-0.012** (0.006)
Female * Share of women in committee			-0.050** (0.025)			-0.010 (0.022)
<i>Controls:</i>						
Candidate characteristics	No	Yes	Yes	No	Yes	Yes
Adj. R-squared	0.081	0.303	0.303	0.006	0.040	0.040
Number of observations	69020	69020	69020	31243	31243	31243

Notes: OLS estimates. All regressions include exam fixed effects. *Candidate characteristics* include all individual predetermined characteristics listed in Table 2. Standard errors are clustered by exam.

* p < 0.10, ** p < 0.05, *** p < 0.01.

Table 5: Randomization check

	1	2	3	4	5	6	7	8	9	10	11
<i>Dependent variable:</i>	All Publ.	Articles	Books	Chapters	Patents	Total AIS	A-journal articles	Coauthors per article	Prop. first-author	Prop. last-author	Age
	Italy										
Female*Share of women in committee	-0.035 (0.085)	0.009 (0.078)	0.071 (0.070)	0.082 (0.065)	-0.046 (0.054)	0.012 (0.068)	-0.077 (0.065)	-0.037 (0.070)	0.102* (0.058)	0.037 (0.097)	0.153* (0.088)
	Spain										
Female*Share of women in committee	0.011 (0.090)	0.037 (0.092)	-0.022 (0.064)	-0.013 (0.065)	-0.073 (0.046)	0.173** (0.080)	0.049 (0.066)	0.120 (0.090)	0.053 (0.093)	0.047 (0.094)	-0.110 (0.102)

Notes: OLS estimates. All regressions include exam fixed effects and the interaction between the variables *Female candidate* and the *Expected share of women in committee*. Standard errors are clustered by exam.

* p < 0.10, ** p < 0.05, *** p < 0.01.

Table 6: The causal impact of committees' gender composition

	1	2	3	4	5	6	7
<i>Dependent variable:</i>	Italy				Spain		
	Qualified		IV	Applied	Qualified		
	OLS	OLS		IV	OLS	OLS	IV
Female candidate	-0.015 (0.009)	0.003 (0.006)	0.009 (0.008)	-0.026*** (0.006)	-.022*** (0.007)	-0.011* (0.007)	-0.011* (0.007)
Female candidate* Share of women in committee	-0.073** (0.036)	-0.107*** (0.026)	-0.132*** (0.035)	-0.025 (0.026)	-0.001 (0.029)	-0.015 (0.028)	-0.016 (0.028)
<i>Controls:</i>							
Candidate characteristics	No	Yes	Yes	Yes	No	Yes	Yes
Number of observations	69020	69020	69020	69020	31243	31243	31243
Mean dep. var. for female candidates	0.35	0.35	0.35	0.83	0.11	0.11	0.11

Notes: All regressions include exam fixed effects and the interaction between the variables *Female candidate* and the *Expected share of women in committee*. *Candidate characteristics* include all individual predetermined characteristics listed in Table 2. Standard errors are clustered by exam.

* p < 0.10, ** p < 0.05, *** p < 0.01.

Table 7: Quality of qualified candidates

	1	2	3	4	5	6	7
<i>Dep. var.:</i>	Publications	Citations	Total AIS	A-journal articles	PhD students advised	PhD thesis committees	Success in future evaluations
A. Italy, before the evaluation							
All	0.001 (0.060)	0.123 (0.102)	-0.117 (0.134)	-0.186 (0.183)			
Women	-0.017 (0.078)	0.148 (0.114)	-0.020 (0.137)	-0.300 (0.234)			
Men	-0.008 (0.083)	0.084 (0.129)	-0.203 (0.187)	-0.071 (0.186)			
B. Spain, before the evaluation							
All	-0.004 (0.142)	0.068 (0.216)	-0.082 (0.237)	-0.200 (0.244)	0.121 (0.135)	-0.143 (0.131)	
Women	0.171 (0.216)	0.446 (0.396)	-0.004 (0.426)	-0.142 (0.357)	0.565** (0.239)	0.052 (0.230)	
Men	-0.149 (0.191)	-0.225 (0.282)	-0.201 (0.292)	-0.218 (0.349)	-0.163 (0.175)	-0.291* (0.168)	
C. Spain, after the evaluation							
All	-0.005 (0.131)	-0.056 (0.211)	-0.092 (0.219)	-0.200 (0.244)	0.169 (0.133)	-0.083 (0.135)	0.040 (0.052)
Women	0.248 (0.220)	-0.009 (0.380)	-0.097 (0.401)	-0.142 (0.357)	0.116 (0.222)	-0.114 (0.243)	0.001 (0.056)
Men	-0.167 (0.181)	-0.131 (0.273)	-0.230 (0.275)	-0.218 (0.349)	0.077 (0.189)	-0.129 (0.184)	0.018 (0.076)

Notes: OLS estimates for the sample of qualified candidates. Each coefficient corresponds to an independent regression for a given sample and dependent variable. In panels A and B the dependent variables are measured at the time of the evaluation. In panel C the dependent variables refer to the output in the five-year period following the evaluation. Success in future evaluations takes value one if a candidate who obtained a qualification for an associate professorship in our sample, qualifies in the evaluation for full professorship by year 2013. The dependent variables in columns 1-6 are normalized to have zero mean and unit variance for candidates within each exam. *Article Influence Score* is only available for candidates in science, technology, engineering, mathematics and medicine. Information on publications in *A-journal articles* is only provided for candidates in social sciences and humanities. All regressions include non-parametric controls for *expected share of women in the committee*, *disciplinary area*rank*, and *age*. Standard errors are clustered by exam.

* p < 0.10, ** p < 0.05, *** p < 0.01.

Table 8: Gender segregation across research networks and subfields

	1	2	3	4	5	6	7	8
	Italy			Spain				
	Colleague	Coauthor	Same subfield	Colleague	Coauthor	PhD Advisor	PhD committee	Research overlap
Female candidate	0.0020*** (0.0003)	0.0005** (0.0002)	0.0162*** (0.0050)	-0.0012 (0.0014)	-0.0003* (0.0002)	-0.0001 (0.0001)	-0.0010*** (0.0003)	0.0052* (0.0027)
Female evaluator	0.0033*** (0.0008)	-0.0004 (0.0004)	0.0325*** (0.0123)	0.0006 (0.0014)	-0.0015*** (0.0002)	-0.0013*** (0.0002)	-0.0047*** (0.0006)	-0.0106*** (0.0016)
Female candidate*	0.0030***	0.0022***	0.0167**	0.0043***	0.0010**	0.0005***	0.0013***	0.0040*
Female evaluator	(0.0007)	(0.0004)	(0.0070)	(0.0016)	(0.0002)	(0.0002)	(0.0005)	(0.0020)
Constant	0.0210*** (0.0002)	0.0112*** (0.0001)	0.4753*** (0.0032)	0.0453*** (0.0001)	0.0045*** (0.0000)	0.0025*** (0.0000)	0.0142*** (0.0002)	0.2010*** (0.0010)
Observations	3,143,428	3,143,428	1,679,376	5,445,067	5,445,067	5,445,067	5,445,067	4,711,621

Notes: OLS estimates. The number of observations corresponds to the number of possible pairs between candidates and eligible evaluators with non-missing information in a given exam. All regressions include exam fixed effects. Standard errors are clustered by candidate.

* p < 0.10, ** p < 0.05, *** p < 0.01.

Table 9: Connections and research similarity

	1	2	3	4	5	6	7	8
	Italy				Spain			
Female candidate	0.009 (0.008)	0.006 (0.008)	-0.007 (0.009)	-0.008 (0.009)	-0.011* (0.007)	-0.011 (0.007)	-0.011 (0.008)	-0.012 (0.008)
Female candidate * Share of female evaluators	-0.132*** (0.035)	-0.127*** (0.034)	-0.070 (0.047)	-0.069 (0.046)	-0.016 (0.028)	-0.019 (0.028)	-0.015 (0.035)	-0.020 (0.035)
<i>Connections in committee:</i>								
Colleagues		0.180*** (0.037)		0.181*** (0.044)		0.318*** (0.031)		0.317*** (0.031)
Coauthors		0.251*** (0.048)		0.221*** (0.054)		0.872*** (0.140)		0.837*** (0.142)
PhD advisors						0.633*** (0.107)		0.570*** (0.115)
PhD thesis committee						0.173*** (0.037)		0.162*** (0.038)
<i>Research similarity:</i>								
Same subfield				0.049 (0.032)				
Overlap in research interests								0.156*** (0.041)
<i>Controls:</i>								
Expected connections		Yes		Yes		Yes		Yes
Expected same subfield				Yes				
Expected overlap in research interests								Yes
Number of observations	69020	69020	35831	35831	31243	31243	27998	27998

Notes: The final share of female evaluators is instrumented by the initial share of female evaluators in the committee. All regressions include exam fixed-effects, an interaction between *Female candidate* and the *Expected share of women in committee*, and controls for all individual predetermined characteristics listed in Table 2. Connection variables are measured in shares. *PhD thesis committee* refers to candidates and evaluators who have been members of the same doctoral thesis committee. *Same subfield* is the share of evaluators who belong to the same subfield (*settore scientifico disciplinario*) as the candidate. *Overlap in research interests* is based on evaluators' and candidates' participation in doctoral thesis committees, which are classified in 2,000 different subfields (see more details in Data section). *Expected connections* is a vector including the expected share in the committee of *colleagues*, *coauthors*, *advisors* and *PhD thesis committee*. Standard errors are clustered by exam.

* p < 0.10, ** p < 0.05, *** p < 0.01.

Table 10: Heterogeneity analysis

	1	2	3	4
	Italy		Spain	
Research overlap	<i>> median</i>	<i>< median</i>	<i>> median</i>	<i>< median</i>
	0.007 (0.046)	-0.193*** (0.067)	0.063 (0.048)	-0.110** (0.043)
Discipline	SSH	STEMM	SSH	STEMM
	-0.117** (0.053)	-0.135*** (0.039)	-0.026 (0.039)	0.003 (0.041)
Feminization of field	<i>> median</i>	<i>< median</i>	<i>> median</i>	<i>< median</i>
	-0.152*** (0.042)	-0.077 (0.056)	-0.018 (0.040)	-0.016 (0.037)
Level of promotion	FP	AP	FP	AP
	-0.107* (0.058)	-0.144*** (0.038)	0.121** (0.054)	-0.072** (0.032)

Notes: IV estimates. The dependent variable is a dummy variable that takes value one if the candidate qualified. Each coefficient corresponds to an independent regression for the corresponding sample. SSH stands for social sciences and humanities, and STEMM for science, technology, engineering, mathematics and medicine. The feminization of the field is measured by the proportion of women among full professors in the discipline. FP and AP stand, respectively, for full and associate professors. Standard errors are clustered by exam.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

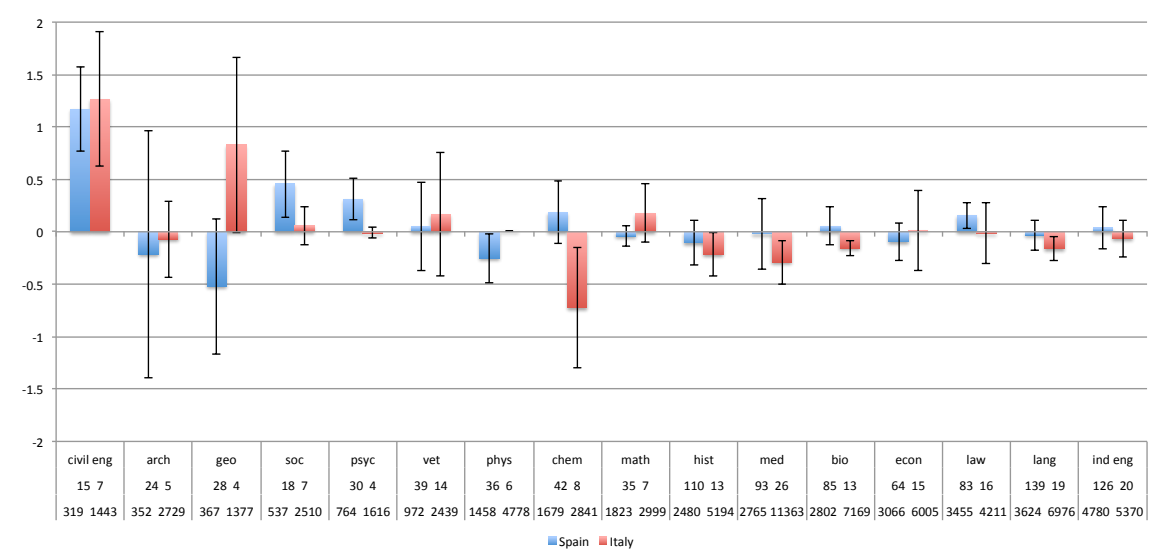
Table 11: Individual voting

	1	2
Female evaluator	0.001 (0.010)	-0.001 (0.014)
Female candidate	-	0.004 (0.007)
Female candidate * Female evaluator	0.006 (0.006)	0.003 (0.013)
Female candidate * Share of women in committee		-0.048* (0.028)
Controls:		
Application FE	Yes	No
Adj. R-squared	0.846	0.411
N	279,427	279,427

Notes: OLS estimates. All regressions include as controls exam fixed-effects and individual predetermined characteristics listed in Table 2. The regression reported in column 2 also includes controls (non-reported) for *Female evaluator*Share of women in committee* and *Female candidate*Female evaluator*Share of women in committee*. Standard errors are clustered by exam.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: The causal impact of committees gender composition, by disciplinary group



Note: The figure reports the causal estimates of the effect of a higher proportion of women among evaluators on the relative success rate of female candidates in the corresponding disciplinary group and country. The confidence intervals are based on committee-clustered standard errors and they are not adjusted for multiple comparisons. At the bottom of the figure, the number of committees and the number of candidates in a corresponding group are shown. The disciplinary groups are sorted according to the number of applicants in each group in Spain.

Appendix A. Data

The data on the participants in Italian evaluations, including the CV of all eligible evaluators and all candidates, was available at the website of the Italian Ministry of Higher Education and Research. We extracted all the individual characteristics that we use in the analysis from these CVs.

The data on the participants in Spanish evaluations was collected from different sources, including the Spanish Ministry of Research and Science, Thomson Reuters (ISI) Web of Knowledge, the database of publications in Spanish language Dialnet, the European Patent Office and TESEO database on doctoral dissertations.²⁴

Publications indexed in above sources are matched to the list of professors in Spain based on individuals' names and field of research. This process suffers from an important problem with homonymity since there are lots of common surnames in Spain. In addition to this, bibliographic databases often incompletely record authors' names (this especially concerns the data on publications before 2010 in the Web of Knowledge). Facing the choice between minimizing the number of false positives or the number of false negatives, we generally preferred the former. This means that, on the one hand, the individuals are authors of the outcomes assigned. On the other hand, we are unable to assign research outputs that have an incomplete record of authors' names.

Below we describe in detail the process of data collection in the case of Spain.

Spanish Ministry of Research and Science The Spanish system of centralized examinations known as 'habilitación' was in place between 2002 and 2006. In total, 1,016 exams took place, around five per discipline. We restrict the sample in several ways. We exclude exams where the number of available positions was larger or equal than the number of candidates (two exams, both in Basque Philology) and disciplines where the number of potential evaluators was not large enough to form a committee (55 exams).²⁵ The final database includes 967 exams.

²⁴The would like to thank Stéphane Maraut and Catalina Martinez for kindly sharing the data on academic inventors who have patented their inventions in the European Patent Office. See Maraut and Martínez (2014) for the description of how the patent data was collected and matched to professors.

²⁵In these cases, unfilled seats in the committee were filled with professors from related disciplines.

Information on candidates' and evaluators' first name, last name, tenure and ID number was retrieved from the website of the Ministry of Research and Science in July 2009 (<http://www.micinn.es>). We used first name information in order to identify gender. In a few cases where it was not possible to assign gender based on first name, we searched online for any personal picture or document that would make it possible to assign gender.

The actual age of individuals is not observable. Instead, we exploit the fact that Spanish ID numbers contain information on their issue date to construct a proxy for the age of native individuals on the basis of his/her national ID number. In Spain, police stations are given a range of ID numbers which are assigned to individuals in a sequential manner. Since it is compulsory for all Spaniards to have an ID number by age 14, two Spaniards with similar ID numbers are likely to be of the same age (and geographical origin).²⁶ In order to perform the assignment, we first use registry information on the date of birth and ID numbers of 1.8 million individuals in order to create a correspondence table which assigns year of birth to the first four digits of ID number (ranges of 10,000 numbers). To test the precision of this correspondence, we apply it to a publicly available list of 3,000 court clerks, which contains both the ID number and the date of birth. In 95% of the cases the assigned age is within a three-year interval of the actual age. In order to minimize potential errors, whenever our age proxy indicated that a candidate for an associate professorship is less than 27 years old and a candidate for full professorship is less than 35 years old, we assign age a missing value. This proxy is also not defined for non-Spaniards (less than 1% of the sample). We imputed the missing age with the average age of individuals at the same discipline and rank (around 5% of the sample).

In 2006 the system of *habilitación* was replaced by a system known as *acreditación*, which is still in place. Under the *acreditación* system applicants aspiring for promotion

²⁶There are a number of exceptions. For instance, this methodology will fail to identify the age of individuals who obtained their nationality when they were older than 14. Nevertheless, immigration was a rare phenomenon in Spain until the late 1990s. Additionally, some parents may have their children obtain an ID number before they are 14. This may be the case particularly after Spain entered in the mid 90s the Schengen zone and IDs became a valid documentation to travel to a number of European countries.

are also required to be approved by a national review committee. These committees evaluate candidacies on a monthly basis and their decisions are published in the Official State Bulletin. We collected information on the identity of all candidates that qualified for a FP position before September 2013.

The Ministry provides information on affiliation and on tenure in the position for eligible evaluators. Given that most candidates to full professor positions are eligible evaluators themselves in exams to associate professor positions, it is possible to obtain their affiliation by matching the list of eligible evaluators with the list of candidates. Using this procedure, we were able to obtain the information on affiliation for 93% of candidates to full professor positions. We obtained the information on affiliation for the remaining 7% of candidates from the State Official Bulletin or directly from professors' CVs that can be found online.

ISI Web of Knowledge We also collected information on the research output of eligible evaluators and candidates from the ISI Web of Knowledge.²⁷

Information on scientific publications comes from the Thomson Reuters ISI Web of Knowledge (WoK). We consider publications published since 1972 by authors based in Spain, as well as the number of citations received by these publications before July 2009. The WoK database includes over 10,000 high-impact journals in the categories of Science, Engineering, Medicine and Social Sciences, as well as international proceedings coverage for over 110,000 conferences. For the purpose of this analysis, we considered all articles, reviews, notes and proceedings.

The assignment of articles to professors is non-trivial. For each publication and author, WoS provides information on his/her surname and on his/her initial. In Spain, some surnames are very common (e.g., Garcia, Fernandez, Gonzalez), and this may create problems with homonymity. Moreover, unlike most other countries, individuals are assigned two surnames (paternal and maternal) and sometimes also several first names. When Spanish authors sign a paper they may do it with only their paternal or

²⁷We are grateful to the *Fundación Española para la Ciencia y la Tecnología* for providing us with access to the data.

with their maternal surname, or they may hyphenate the two surnames. Authors may also sign using their first name, their middle name, or both.

We use the following matching procedure in order to deal with the above problems. First, we assign all publications and all professors in our sample to a broad disciplinary category. In order to attribute comparable disciplinary categories for publications and individuals, we aggregate disciplines defined by the Spanish Ministry and ISI disciplinary areas into the following categories: Agriculture; Chemistry; Biology; Geology; Physics; Mathematics and Computer Science; Engineering; Medicine, Veterinary and Pharmacology; Economics and Management; Psychology, Sociology and Political Science.²⁸ Second, in each broad disciplinary category we match publications with individuals in our database using the information on their surnames and initials.

Specifically, the publication is assigned to a professor in the list of eligible evaluators if it is in the same disciplinary category as the professor, and the author's surname and initial, as reported by ISI, coincide (i) with the first surname and the first name's initial of the professor, (ii) with the last surname and the first initial, (iii) with the first surname hyphenated with the second surname and the first initial. We also repeat stages (i) through (iii) substituting the first initial with the middle-name initial. If a given publication can be assigned to more than one possible match, the value of this publication is divided by the number of such possible matches.

Given that the propensity to publish differs substantially across the disciplines, we normalize the number of individual's publications to have zero mean and unit standard deviation among applicants to the same exam and among eligible evaluators of a given category in a given exam. The number of citations of each publication depends on the time elapsed between the publication date and the date when the number of received citations is observed. Therefore, we first normalize the number of citations that each publication receives by subtracting the average number of citations received by Spanish-authored articles published in the corresponding ISI disciplinary

²⁸In practice, apart from the case of journals *Science* and *Nature*, the ISI scientific categories are assigned to journals, not publications. In very rare cases a publication happened to be assigned to more than one broad disciplinary group.

area in the same year and then dividing by the corresponding standard deviation. Next, for each individual in our database we calculate the average number of citations per publication. For individuals who have no ISI publications, this variable takes the minimum value in the corresponding discipline. Finally, similarly to the number of publications, we normalize the number of individual's citations per publication to have zero mean and unit standard deviation among applicants to the same exam and among eligible evaluators of a given category in a given exam.

Dialnet Dialnet (<http://dialnet.unirioja.es>) is an open access bibliographic index created by the University of Rioja. It contains information on more than 8,000 journals and more than 3,5 million documents in Hispanic languages, including articles published in scientific journals, collective works and books. The database mainly covers publications in social sciences and humanities. Dialnet provides (in most cases) systematized information on individual authors' first name, paternal surname, maternal surname and affiliation, thus limiting potential concerns about homonymity.

We collected information on publications in Dialnet. Due to its lack of representativeness, we did not consider publications in Science and Engineering. We also excluded publications that appear in ISI Web of Science. We also restricted the set of journals considered to those which satisfy certain minimum research quality requirements (categories A, B or C) as established by the Integrated Scientific Journals Classification (CIRC) (Torres-Salinas et al. 2010). Similarly, we considered only books and collective volumes that are published by publishers that satisfy a minimum quality requirement. In particular, we used the EPUC-CSIC publisher list, which summarizes the names of the main publishers in social sciences and humanities in Spain and abroad (Giménez-Toledo, Tejada-Artigas and Mañana-Rodríguez 2012). Publications that have been excluded from our study are mainly publications in working paper series, non-refereed journals and volumes published by local universities (around 30%).

Teseo database on doctoral dissertations Since 1977, PhD candidates in Spanish universities have registered their dissertation in the database TESEO, which is run by

the Ministry of Education. We retrieved all the information available in this database from the website <https://www.educacion.gob.es/teseo> in May 2011. While registration is compulsory, according to Fuentes and Arguimbau (2010) TESEO includes information on approximately 90% of all dissertations read in Spain during this period. We observe information on 151,483 dissertations. TESEO provides the identity and affiliation of dissertations' authors, advisors and committee members. Approximately 40% of dissertations are female authored. Female supervisors are scarce and represent only 18% of the total. While 58% of the students they supervise are female, in the case of male advisors, 61% of their students are male.

We match TESEO data with the list of candidates and evaluators. In exams to full professor positions we are able to find the dissertation of 71% of candidates and 41% of evaluators. In exams to associate professor positions we observe the dissertation of 83% of candidates and 70% of evaluators. Missing information may be due to the fact that individuals (i) did their PhD abroad, (ii) defended their dissertation before 1977, (iii) there are spelling mistakes, (iv) the dissertation was not included in TESEO for unknown reasons (approximately 10% of all dissertations), or (v) there was a problem with homonymity (in our dataset 0.1% of individuals share the same name, middle name, paternal surname and maternal surname).

Each thesis has been classified by its author using the Unesco International Standard Nomenclature for Fields of Science and Technology. This is a system developed by Unesco that includes more than two thousand six-digits categories.²⁹ 80% of dissertations provide this information. Approximately half of the authors select one six-digit category, 35% select two categories, and 15% select three or more categories. There are on average around one hundred dissertations per category. We use this information to construct a measure of individuals' research interests. In particular, we take into account every dissertation where an individual appears as an advisor, committee member or author. We were able to obtain information on the research interests of 98% candidates to full professor positions, 94% of candidates to full professor positions,

²⁹Available at <http://unesdoc.unesco.org/images/0008/000829/082946eb.pdf>

98% of eligible full professors and 96% of eligible associate professors.

Appendix B. Nonlinearities

The effect of the gender composition of committee on the relative success rate of females may be non-linear for a number of reasons. First, the presence of a woman in the committee may affect the voting behavior of male evaluators (see section 4.5). If this is the case, the transition from zero to one female evaluator in the committee may have a different effect than the transition from one to two female evaluators, or from two to three female evaluators. Second, decisions in the committee are taken on a (qualified) majority basis. Therefore, having a committee where the (qualifying) majority of members are female might have a particularly strong effect.

In order to correctly identify the potential existence of nonlinear effects, it is necessary to control for the probability that a given number of women is assigned to the committee. We consider the following model:

$$y_{ie} = \beta_0 + \beta_1 Female_i + \sum_k \gamma_k Female_i D_{ke} \\ + \sum_k \delta_k Female_i D_{ke}^{expected} + \mathbf{X}_i \beta_2 + \mathbf{Z}_i \beta_3 + \mu_e + \epsilon_{ie}$$

where D_{ke} is a dummy variable that takes value one if the number of female evaluators in committee e is equal to k and $D_{ke}^{expected}$ is the probability that exactly k female evaluators are assigned to a given committee. For Spanish evaluations, we directly compute these probabilities using information on the gender mix of the pool of eligible evaluators. For the Italian case, the direct computation is more complicated, since the assignment procedure required no more than one committee member from each university. Instead, we compute these probabilities using the outcomes of 1000 simulated random draws, which account for the restrictions on the randomization.

Committees rarely included more than three women. Therefore, we only analyze the effect of having one, two, and three or more female evaluators. The estimation results are presented in Table A3. Overall, the linearity of the effect of committees' gender composition cannot be rejected by the data.

Appendix C. The effect of connections, by gender of evaluators and candidates

As we have seen in section 4.4.2, there is significant gender segregation across networks but it is not strong enough to manifest itself as apparent preference for candidates of the same sex as evaluators. One might be interested whether this is in part due to the differential impact of same-sex and opposite-sex connections upon female and male candidates. In Table A4, we explore this issue in more detail. We do not observe any differential effect of strong connections (coauthors, colleagues and, in the case of Spain, advisors) for female and male candidates. We also do not observe that the female and male connections have a differential effect on candidates' success.

Appendix Tables

Table A1: Literature review

Paper	Type of evaluation	Field	Empirical method	Applications	Results
Broder (1993)	Grant applications	Economics	Application fixed effects	1,479	Opposite-sex preference
Steinpreis, Anders and Ritzke (1999)	Job applicants and tenure candidates	Psychology	Randomized field experiment	238	No significant difference
Jayasinghe, Marsh and Bond (2003)	Grant applications	Several	Application fixed effects	2331	No significant difference
Ellemers et al. (2004)	Work commitment of students	Several	Identification based on observables	212	Opposite-sex preference
Moss-Racusin et al. (2012)	Laboratory manager position	Life Sciences	Randomized field experiment	127	No significant difference
Abrevaya and Hamermesh (2012)	Paper submitted for publication	Economics	Application fixed effects	2,940	No significant difference
Casadevall and Handelsman (2013)	Selection of conference speakers	Microbiology	Identification based on observables	1,845	Same-sex preference
De Paola and Scoppa (2015)	Job applicants	Economics and Chemistry	Identification based on observables	2,279	Same-sex preference
Williams and Ceci (2015)	Job applicants	Several	Randomized field experiment	873	No significant difference

Table A2: Main features of the evaluation systems in Italy and Spain

	Italy, Abilitazione Scientifica Nazionale, 2012-2014	Spain, Habilitación, 2002-2006
Eligibility requirement for candidates	None	None
Size of evaluation committees	5 evaluators	7 evaluators
Assignment to committees	Based on a random draw	Based on a random draw
Composition of committees	4 full professors based in Italian universities, 1 professor based abroad	In full professor exams, 7 full professors based in Spanish universities or public research centers. In associate professor exams, 3 full professors and 4 associate professors.
Constraints on randomization	No university can have more than one evaluator within a single committee.	Only one non-university researcher is allowed to be selected as a member of the committee for a given exam. Similarly, only one emeritus professor is allowed to be selected as a member of a given committee.
Minimum research quality requirement for evaluators	In STEMM disciplines, eligible professors should be above the median in their category and field in at least two of the following dimensions: (i) the number of articles published in scientific journals, (ii) the number of citations, (iii) and the H-index. In SSH disciplines, they should be above the median in at least one of the following dimensions: (i) the number of articles published in high quality scientific journals (so-called A-journals), (ii) the overall number of articles published in any scientific journals and book chapters, and (iii) the number of published books.	Eligible associate professors should have one <i>sexenio</i> and eligible full professors should have two <i>sexenios</i> . <i>Sexenios</i> are granted by the Spanish education authority on the basis of applicants' research output in any non-interrupted period of a maximum of six years.
Inclusion in the pool of eligible evaluators	Voluntary	Compulsory
Substitution of resigned evaluators	Based on a random draw	Based on a random draw
Voting rule	Qualified majority of 4	Simple majority
Number of qualifications granted by the committee	Unlimited	Limited by the number of available positions at the university level
Validity of a positive qualification	4 years (later extended to 6 years)	Unlimited
Penalization for a negative evaluation	2 years application ban	None
Application withdrawal	Up until two weeks after the evaluation criteria are publicized	Candidates can drop out from the process at any time
Evaluation	Evaluations are based solely on the material provided in candidates' application packages, consisting of CVs and selected publications.	Oral exams to full professor positions have two qualifying stages. In the first stage, candidates present their CVs. In the second stage, candidates present a piece of their research work. Exams to associate professor, in addition to these two stages, have an intermediate stage where candidates give a lecture on a topic randomly chosen from a syllabus proposed by the candidate.
Degree of transparency	The lists of potential and actual evaluators and candidates, as well as the lists of qualified candidates, are published online. Furthermore, the CVs of all participants and individual evaluation reports are published online. The evaluation agency also collects and publicizes information on the bibliometric indicators of candidates.	The lists of potential and actual evaluators and candidates, as well as the lists of qualified candidates, are published online.

Table A3: Nonlinearities

	1	2
	Italy	Spain
Female	0.001 (0.007)	-0.012 (0.007)
Female * 1 female evaluator	-0.018 (0.012)	-0.001 (0.011)
Female * 2 female evaluators	-0.037*** (0.012)	-0.006 (0.013)
Female * 3 or more female evaluators	-0.080*** (0.021)	-0.004 (0.014)
Number of observations	69020	31243

Notes: IV estimates. All regressions include as controls exams fixed-effects, the number of female evaluators in the committee, individual predetermined characteristics, and the expected probabilities to have 1, 2, and 3 or more female evaluators. Standard errors are clustered by exam.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: The effect of strong connections, by candidate and evaluator gender

	1	2
	Italy	Spain
Female candidate	0.007 (0.008)	-0.012* (0.007)
Female candidate * Share of female evaluators	-0.134*** (0.035)	-0.011 (0.028)
Share of connections in committee	0.216*** (0.041)	0.429*** (0.038)
Female candidate * Share of connections in committee	-0.002 (0.066)	0.017 (0.060)
Share of female connections in committee	-0.031 (0.084)	-0.039 (0.101)
Female candidate * Share of female connections in committee	0.164 (0.124)	-0.077 (0.145)
Number of observations	69020	31243

Notes: IV estimates. All regressions include exam fixed-effects, individual predetermined characteristics, *Female candidate * Expected share of women in committee*, *Expected connections in committee*, *Female candidate * Expected connections in committee*, *Expected female connections in committee* and *Female candidate * Expected female connections in committee*. Standard errors are clustered by exam.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Achieving Speaker Gender Equity at the American Society for Microbiology General Meeting

Arturo Casadevall

Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

ABSTRACT In 2015, the American Society for Microbiology (ASM) General Meeting essentially achieved gender equity, with 48.5% of the oral presentations being given by women. The mechanisms associated with increased female participation were (i) making the Program Committee aware of gender statistics, (ii) increasing female representation among session convener teams, and (iii) direct instruction to try to avoid all-male sessions. The experience with the ASM General Meeting shows that it is possible to increase the participation of female speakers in a relatively short time and suggests concrete steps that may be taken to achieve this at other meetings.

IMPORTANCE Public speaking is very important for academic advancement in science. Historically women have been underrepresented as speakers in many scientific meetings. This article describes concrete steps that were associated with achieving gender equity at a major meeting.

Received 9 July 2015 Accepted 15 July 2015 Published 4 August 2015

Citation Casadevall A. 2015. Achieving speaker gender equity at the American Society for Microbiology General Meeting. *mBio* 6(4):e01146-15. doi:10.1128/mBio.01146-15.

Editor Michael J. Imperiale, University of Michigan

Copyright © 2015 Casadevall. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Arturo Casadevall, Acasade1@jhu.edu.

Invitations to speak at major meetings are prized by scientists because they provide visibility and the ability to present their work efficiently to an audience of peers. Speaking invitations are used by faculty promotion and tenure committees as evidence of external recognition and thus can be critical to academic and professional advancement. Studies have shown that women's participation at meetings is often underrepresented relative to their numbers (1, 2). In fact, women lag in a variety of academic areas despite increases in the percentage of women scientists (3–6). Given the importance of oral presentations at scientific meetings and academic advancement, a colleague and I recently analyzed the role of convener gender in symposia in two meetings from 2011 to 2013 and noted that the presence of at least one female convener was associated with a marked increase in female speaker participation and a reduction in the probability that the session was all male (7). This communication describes the effect of making that information known to the American Society for Microbiology (ASM) General Meeting (GM) Program Committee in the years 2014 and 2015 and discusses interventions that were associated with achieving gender parity in 2015 (Fig. 1).

The ASM GM is planned by a Program Committee that meets in Washington, DC, in the summer before the meeting. The committee is composed of scientists who represent the wide breadth of subdisciplines that participate in the GM, ranging from microbial pathogenesis to microbial ecology and basic microbial physiology and genetics. The format for the organization of the GM has not been changed since the meeting was reorganized in 2011 (8), and the process was described in detail in a prior publication (7). Basically, the Program Committee designs symposia and assigns shepherds who are members of the committee, who in turn choose conveners who select the speakers. Hence, the convener

team has great latitude in speaker selection, with the shepherds overseeing the process and occasionally providing input into session content.

In the summer of 2013, the results of the GM gender participation analysis for the years 2011 to 2013, subsequently published in *mBio* (7), were available and were presented to the Program Committee planning the 2014 meeting in Washington, DC. Two points were made: (i) that female speakers were underrepresented relative to the participation of females in the meeting, with too many all-male sessions, and (ii) that the presence of at least one female convener was associated with a 72% increase in female participation in those sessions and a 70% reduction in the likelihood of an all-male session. The simple intervention of presenting the data from prior meetings was associated with an increase in convener teams comprising at least one woman and an increase in female speaker participation at the 2014 meeting to 43% from an average of 29.6% for the prior 3 years but had no effect on the number of all-male sessions (Fig. 1). In the summer of 2014, the gender statistics from the 2014 meeting (which had occurred 1 month earlier) were again presented to the Program Committee planning the 2015 meeting, together with an analysis of trends in recent years. However, on that occasion, the committee was also instructed verbally to “do better” with regard to gender balance and to avoid all-male sessions, except under extraordinary circumstances. The PowerPoint slide used to instruct the committee during the presentation is shown in Fig. 2. The 2015 ASM GM essentially achieved gender equity among speakers, with 48.5% being women. As a result of the increases in the percentage of women speaking in the 2014 and 2015 meetings, approximately 100 more women presented in those two meetings than the number expected on the basis of the 2011 to 2013 meeting averages.

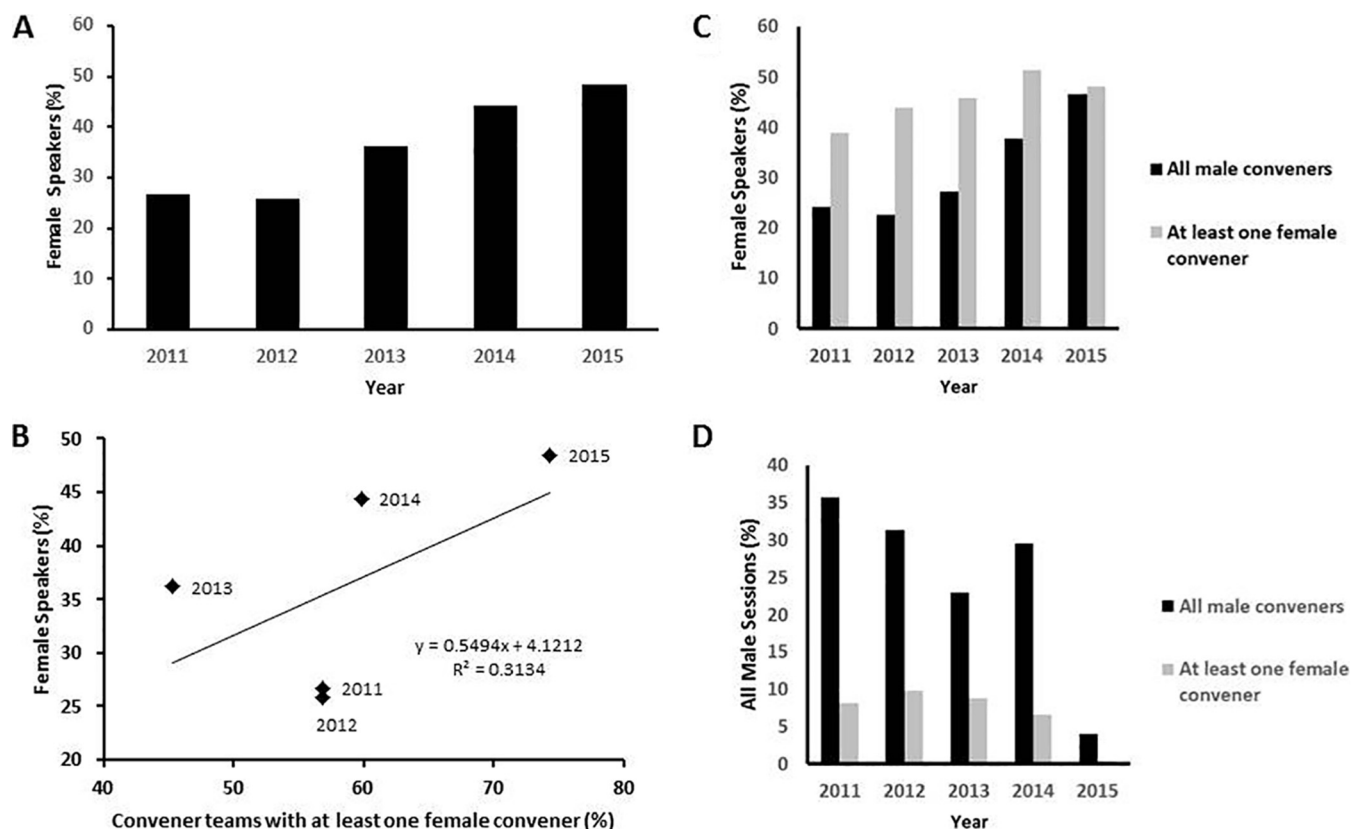


FIG 1 Panel A shows the percentages of female speakers for five consecutive ASM GMs in the years 2011 to 2015. Panel B shows a plot of the percentage of session convener teams with a least one female convener versus the percentage of women speakers with the year next to the data point. Panel C shows the percentages of female speakers in sessions organized by all-male conveners and those organized by teams with at least one female convener. Panel D shows the percentages of all-male sessions among sessions organized by all-male convener teams and teams with at least one female convener. Data for the years 2011 to 2013 were published previously (7) and are shown here for comparison with those for the years 2014 and 2015.

Two other outcomes were striking. First, the percentage of all-male speaker sessions was reduced tremendously, and second, the gender difference among conveners disappeared, such that all-male convener teams were equally as likely to invite female speakers as those that included a female convener (Fig. 1). A positive

correlation between the presence of at least one woman on the convener team and the percentage of women speakers was observed when the data from the five meetings were combined (Fig. 1).

These results establish that it is possible to achieve gender equity among speakers in a major scientific meeting in a reasonably short time frame. Although increases in female speaker participation followed two interventions with the Program Committee involving first the presentation of gender statistics and second a verbal plea to do better with gender balance, I caution that this report merely establishes an association between these interventions and increased female participation and remind the reader that association is not causation. Although it is conceivable that the increase in female participation was a result of the inclusion of more topics where female scientists were better represented, I do not think this explains the data since the overall mix of scientific topics at the ASM GM is fairly constant year after year. Another possibility is that women have more women in their scientific networks and that their presence in convener teams resulted in more suggestions for women speakers. In this regard, the establishment of networks can contribute to success for women in science (9).

It is likely that making the committee aware of the gender statistics influenced their decisions, increasing the number of female conveners and potentially encouraging shepherds to be more pro-

CHARGE TO THE COMMITTEE

- **YOU HAVE MADE SIGNIFICANT PROGRESS BUT YOU NEED TO DO BETTER**
- **DISCUSS THESE STATISTICS WITH THE CONVENERS YOU SELECT** (CAN SHARE CASADEVALL & HANDELSMAN *Mbio* 2014 PAPER)
- **AIM FOR GENDER DIVERSITY AT THE CONVENER LEVEL: DATA SUGGESTS THAT HAVING A FEMALE CONVENER INCREASES FEMALE PARTICIPATION AND REDUCES ALL MALE ROSTERS**
- **IF YOU SEE A PROPOSED ALL MALE ROSTER PLEASE DEMAND AN EXPLANATION – IT IS OK TO INTERVENE!**
- **LET'S WORK TOGETHER AND USE DATA TO REDUCE CONSCIOUS AND UNCONCIOUS BIASES**

FIG 2 Slide used to charge the Program Committee during the summer 2014 meeting.

active in enhancing female participation in their assigned sessions. Particularly noteworthy was the observation that the difference in female speakers associated with convener teams that included at least one female convener disappeared in 2015. One explanation for this effect may be that all-male convener teams responded to the data presented to the committee and published previously (7) and decided to actively recruit more female speakers. Alternatively, it is possible that all-male teams felt that they were under increased scrutiny and increased their selection of women speakers. Whatever the explanation, the outcome resulted in significantly increased numbers of women speakers.

Among ASM members, women now constitute the majority of the students and postdocs (7) and therefore represent the future of the society. Thus, it is important to convey the message to the next generation that in this field there is no “glass ceiling” with regard to gender. Since selection as a speaker is an important form of recognition that can carry tangible benefits with regard to career development, the achievement of gender equity at the ASM GM will, I hope, convey the message that both genders are welcomed into the accomplished ranks of the Society. In this regard, all-male sessions are potentially problematic, as they could be interpreted as indicative of areas with poor gender balance that are perhaps not as welcoming to women. The near eradication of all-male sessions at the 2015 ASM GM shows that it is possible to effect change in this type of session format and thus avoid any subtle negative messages to female scientists in training and younger faculty members.

Our results suggest three strategies that may be used to increase the participation of women in meetings where better gender balance is needed and desired. First, obtaining gender data from prior meetings and presenting them to the Program Committee can increase awareness of inequities in gender balance. Second, the increased representation of women among those who select speakers was associated with increased female speaker participation. Third, direct instruction to the committee can help focus the group on reducing inequities in gender balance.

ACKNOWLEDGMENTS

I thank Janet Mitchell for her efforts in compiling the ASM gender data and organizing them for analysis. Credit for increasing female participation belongs to the Program Committee, which selected conveners, suggested speakers, and worked to develop symposia. Special thanks to program chair elect Nicole Dubilier and to ASM Meetings Board Chair David Hooper for their active support in this effort.

I served as Chair of the ASM GM Program Committee for the years 2014 and 2015.

REFERENCES

1. Simon JL, Morris EK, Smith NG. 2007. Trends in women's participation at the meetings of the Association for Behavior Analysis: 1975–2005. *Behav Anal* 30:181–196.
2. Isbell LA, Young TP, Harcourt AH. 2012. Stag parties linger: continued gender bias in a female-rich scientific discipline. *PLoS One* 7:e49682. <http://dx.doi.org/10.1371/journal.pone.0049682>.
3. West JD, Jacquet J, King MM, Correll SJ, Bergstrom CT. 2013. The role of gender in scholarly authorship. *PLoS One* 8:e66212. <http://dx.doi.org/10.1371/journal.pone.0066212>.
4. Cho AH, Johnson SA, Schuman CE, Adler JM, Gonzalez O, Graves SJ, Huebner JR, Marchant DB, Rifai SW, Skinner I, Bruna EM. 2014. Women are underrepresented on the editorial boards of journals in environmental biology and natural resource management. *PeerJ* 2:e542. <http://dx.doi.org/10.7717/peerj.542>.
5. Rotbart HA, McMillen D, Taussig H, Daniels SR. 2012. Assessing gender equity in a large academic department of pediatrics. *Acad Med* 87:98–104. <http://dx.doi.org/10.1097/ACM.0b013e31823be028>.
6. Addessi E, Borgi M, Palagi E. 2012. Is primatology an equal-opportunity discipline? *PLoS One* 7:e30458. <http://dx.doi.org/10.1371/journal.pone.0030458>.
7. Casadevall A, Handelsman J. 2014. The presence of female conveners correlates with a higher proportion of female speakers at scientific symposia. *mBio* 5(1):e00846–13. <http://dx.doi.org/10.1128/mBio.00846-13>.
8. Miller JF, McFall-Ngai M, Casadevall A. 2010. A new design for the ASM General Meeting. *mBio* 1(5):e00240–10. <http://dx.doi.org/10.1128/mBio.00240-10>.
9. Stenken JA, Zajicek AM. 2010. The importance of asking, mentoring and building networks for academic career success—a personal and social science perspective. *Anal Bioanal Chem* 396:541–546. <http://dx.doi.org/10.1007/s00216-009-3275-x>.

FACULTY INSIGHTS ON EDUCATIONAL DIVERSITY

Meera E. Deo*

Twice in the past two years, the U.S. Supreme Court has approved educational diversity as a compelling state interest that justifies the use of race in higher education admissions decisions. Nevertheless, it remains on somewhat shaky ground. Over the past decade, the Court has emphasized that its acceptance of diversity stems from the expectation that a diverse student body will enhance the classroom environment, with students drawing on their diverse backgrounds during classroom conversations that ultimately bring the law to life. Yet, the Court provides no support for its assumption that admitting and enrolling diverse students actually result in these educational benefits. In fact, empirical research on law students indicates that structural diversity (i.e., diversity in numbers) does not lead automatically to interactional diversity (i.e., meaningful interaction among diverse students) or classroom diversity (i.e., meaningful diverse interaction in the classroom specifically); instead, these enhanced classroom experiences depend on adept facilitation by faculty and mutual respect among diverse students.

The Court could draw from a wide body of empirical scholarship with students to better understand the ways in which educational diversity could provide true scholastic and professional benefits. Yet, another group of classroom participants and observers offers even more astute perspectives. Law faculty members have never been asked about their perspectives on educational diversity as part of a formal empirical study, though as the ones facilitating discussion, leading classroom conversations, and instilling a model of respect, they have unique experiences and insights into the possible benefits of educational diversity.

This Article presents findings from the Diversity in Legal Academia (DLA) project, a landmark empirical study of the law faculty experience. DLA findings suggest that law faculty members from all racial/ethnic

* Associate Professor of Law, Thomas Jefferson School of Law. The author thanks the many supporters of the Diversity in Legal Academia (DLA) project, from inception through current dissemination. Kale Sopoaga, Brittany Nobles, Jillian Kates, and Eva Kobi provided outstanding research assistance. Thanks are especially due to the ninety-three law faculty members whose perspectives are included in this Article as DLA findings. This Article is part of a larger symposium publishing articles from the *Critical Race Theory and Empirical Methods Conference* held at Fordham University School of Law in November 2014. For an overview of this symposium, see Kimani Paul-Emile, *Foreword: Critical Race Theory and Empirical Methods Conference*, 83 FORDHAM L. REV. 2953 (2015).

backgrounds not only appreciate the many benefits of diversity, but they also recognize the educational and professional challenges associated with the lack of diversity currently plaguing many law schools. Courts, administrators, and others should rely on these findings to provide additional support for affirmative action through educational diversity, especially to bolster it while it is under attack.

INTRODUCTION.....	3116
I. STRICT SCRUTINY IN THEORY AND PRACTICE	3118
A. <i>The Constitutional Standard</i>	3118
B. <i>Statewide Initiatives and Court Challenges</i>	3119
C. <i>The Current State of Educational Diversity</i>	3122
II. EMPIRICAL SCHOLARSHIP ON EDUCATIONAL DIVERSITY	3124
A. <i>Student Attitudes</i>	3124
B. <i>Faculty Preferences</i>	3126
C. <i>The Diversity in Legal Academia Project</i>	3128
III. EMPIRICAL FINDINGS OF LAW FACULTY SUPPORT FOR DIVERSITY	3133
A. <i>Support for Diversity</i>	3133
B. <i>Benefits of Diversity</i>	3138
1. <i>Diverse Perspectives</i>	3139
2. <i>Personal Context</i>	3142
3. <i>Professional and Career Advantages</i>	3146
C. <i>Missed Opportunities</i>	3147
1. <i>Holes in the Conversation</i>	3147
2. <i>Less Fulfilling Discussions</i>	3149
3. <i>Being a Spokesperson</i>	3150
CONCLUSION	3152

INTRODUCTION

From the days of *Grutter v. Bollinger*,¹ legal scholars as well as those in education, sociology, and related fields have lavished considerable attention on understanding, discussing, and dissecting educational diversity. Legal scholars have examined theoretical bases of the diversity rationale, suggesting that by “drawing on their experiences and contributing their unique viewpoints,” students improve classroom learning.² Social science experts have conducted empirical research on structural, interactional, and classroom diversity, noting that while numeric representation does not lead automatically to meaningful interaction between diverse students in the classroom or elsewhere on campus, it is a prerequisite for those eagerly

1. 539 U.S. 306 (2003).

2. Devon W. Carbado & Mitu Gulati, *What Exactly Is Racial Diversity?*, 91 CALIF. L. REV. 1149, 1158–60 (2003) (reviewing ANDREA GUERRERO, *SILENCE AT BOALT HALL: THE DISMANTLING OF AFFIRMATIVE ACTION* (2002)).

sought-after enhanced classroom conversations where students rely on personal experience to illuminate discussions of substantive law.³

Courts have acknowledged the growing body of scholarship supporting the many benefits of educational diversity. The U.S. Supreme Court has agreed that diversity creates “substantial” benefits for students in the law school context.⁴ These range from “more spirited” classroom discussions⁵ to “the lessening of racial isolation and stereotypes.”⁶ The Court specifically predicts that educational diversity yields an “enhanced classroom dialogue” that is so compelling that universities may pursue them through well-crafted affirmative action policies.⁷

Yet, each of these studies and court opinions relies on only the *student* perspective. Missing from this rich literature are faculty insights into whether educational diversity matters and, if so, how and why.

Ironically, in spite of their empirical silence on this matter until now, faculty members are in the best position to gauge teaching and learning in the classroom and therefore make judgments as to the worth of educational diversity. Because faculty members are tasked with imparting knowledge and facilitating student understanding of legal material, faculty perspectives on student learning in both diverse and non-diverse environments are crucial to forming a true understanding of the value of diversity in higher education.

Educational diversity is the sole surviving non-remedial rationale supporting affirmative action as a compelling state interest.⁸ However, because the Supreme Court has retreated somewhat in its support of educational diversity, the long-term future viability of this rationale is unclear.⁹ This Article draws from a national dataset of law faculty members, ranging from assistant professor to dean emeritus in every region of the United States and from diverse racial/ethnic backgrounds, to explore faculty perspectives on educational diversity. Empirical data show that law faculty members appreciate student diversity, especially because it brings unique perspectives into the classroom, energizing discussions, improving the educational process, and creating benefits for students’ future professional careers. Those who teach in non-diverse environments see the

3. See, e.g., Meera E. Deo, *The Promise of Grutter: Diverse Interactions at the University of Michigan Law School*, 17 MICH. J. RACE & L. 63 (2011); Patricia Gurin, Expert Report, *The Compelling Need for Diversity in Higher Education*, 5 MICH. J. RACE & L. 363 (1999) (prepared for litigation in *Gratz v. Bollinger* and *Grutter*).

4. See *Grutter*, 539 U.S. at 330.

5. See *id.*

6. See *Fisher v. Univ. of Tex. at Austin*, 133 S. Ct. 2411, 2418 (2013).

7. See *id.*

8. Meera E. Deo, *Empirically Derived Compelling State Interests in Affirmative Action Jurisprudence*, 65 HASTINGS L.J. 661, 668 (2014) (“In the context of higher education, educational diversity is the only non-remedial compelling state interest that courts have sanctioned to date.”).

9. See *id.* at 664 (“Yet, diversity has been under attack in past years and faces uncertainty in the future. Justice O’Connor suggested in *Grutter v. Bollinger* that the educational diversity rationale may have a limited shelf life, and *Fisher v. University of Texas* recently narrowed strict scrutiny further.”).

flip side: a narrow set of perspectives, few fulfilling conversations about the context of the law, and alienation of the few students of color in class.

Drawing from this empirical research, this Article proposes that administrators, courts, and policy makers draw from faculty perspectives on educational diversity when discussing the continued importance of this compelling state interest. Adding the faculty perspective to existing scholarship will provide a more complete picture of the benefits of educational diversity.

I. STRICT SCRUTINY IN THEORY AND PRACTICE

A. *The Constitutional Standard*

Parties defending the constitutionality of affirmative action in higher education admissions decisions must show that their policies satisfy “strict scrutiny,” because when government decisions “touch upon an individual’s race or ethnic background, he is entitled to a judicial determination that the burden he is asked to bear on that basis is precisely tailored to serve a compelling governmental interest.”¹⁰ Simply put, the race-conscious policy at issue in each case must both serve a compelling state interest and be narrowly tailored to achieve that interest.¹¹

Interestingly, both prongs are currently in a state of doctrinal flux. Narrow tailoring, the second prong of the strict scrutiny analysis, has become narrower over time.¹² The Court recently asserted that “[e]ducational institutions being sued for using race as a factor in admissions must convince the trial court ‘that it is necessary for a university to use race to achieve the educational benefits of diversity.’”¹³ In other words, “Only if ‘no workable race-neutral alternatives would produce the educational benefits of diversity’ may the university maintain an admissions policy that takes account of race.”¹⁴ This moves the narrow tailoring prong of strict scrutiny into a “least restrictive means” requirement.¹⁵

Educational diversity, the first prong of strict scrutiny, may be nearing its end point as it loses favor with the Court.¹⁶ Litigants have presented both trial and appellate courts with a number of possible compelling state interests over the years, ranging from serving underrepresented

10. See *Fisher*, 133 S. Ct. at 2417 (quoting *Regents of Univ. of Cal. v. Bakke*, 438 U.S. 265, 299 (1978)).

11. *Id.*

12. Deo, *supra* note 8, at 673.

13. *Fisher*, 133 S. Ct. at 2420; see also Rebecca K. Lee, *Fisher v. University of Texas at Austin: Promoting Full Judicial Review and Process in Applying Strict Scrutiny*, 4 HOUS. L. REV. HLR 33 (2013).

14. Deo, *supra* note 8, at 672 (quoting *Fisher*, 133 S. Ct. at 2420).

15. *Id.*

16. See *id.* at 664 (“[D]iversity has been under attack in past years and faces uncertainty in the future. Justice O’Connor suggested in *Grutter v. Bollinger* that the educational diversity rationale may have a limited shelf life, and *Fisher v. University of Texas* recently narrowed strict scrutiny further.”).

communities¹⁷ to addressing widespread societal discrimination.¹⁸ Additional potential compelling state interests, including minimizing racial isolation and diversifying American leadership, have appeared in the academic literature, drawing from empirical data on law students.¹⁹ Yet, the Supreme Court has authorized only one non-remedial compelling state interest as justifying the use of race in admissions decisions: educational diversity.²⁰

B. Statewide Initiatives and Court Challenges

Challenges to affirmative action in the higher education admissions context began with the 1978 case *Regents of University of California v. Bakke*.²¹ In that case, an unsuccessful white applicant to the University of California at Davis Medical School alleged that the school's practice of reserving a certain number of seats for underrepresented applicants violated state and federal law.²² Though no clear consensus emerged, Justice Powell authored the opinion that governed affirmative action policies for the next twenty-five years.²³ In his opinion, Justice Powell determined that institutions of higher learning were required to evaluate all applicants together;²⁴ while schools could not maintain separate applicant pools, they could award a "plus" factor to applicants who they believed would contribute to the diversity of the school.²⁵ He modeled this proposal on the "Harvard Plan," the affirmative action policy in place at that institution.²⁶ Although Justice Powell rejected a number of other possible compelling

17. *Regents of Univ. of Cal. v. Bakke*, 438 U.S. 265, 310 (1978); see also Deo, *supra* note 8, at 669 n.47 ("The [defendant in *Bakke*] also advanced the idea that students of color would graduate and work in underserved populations as a compelling state interest, though the Court determined that there was insufficient evidence in the record to rely on that interest.").

18. See, e.g., *Grutter v. Bollinger*, 539 U.S. 306, 306 (2003); *Wygant v. Jackson Bd. of Educ.*, 476 U.S. 267, 273 (1986). More recently in *Fisher*, the Court reaffirmed that addressing widespread societal discrimination "could not serve as a compelling interest, because a university's 'broad mission [of] education' is incompatible with making the 'judicial, legislative, or administrative findings of constitutional or statutory violations' necessary to justify remedial racial classification." *Fisher*, 133 S. Ct. at 2417 (alteration in original) (quoting *Bakke*, 438 U.S. at 307, 309).

19. Deo, *supra* note 8, at 690–98, 702–05.

20. *Id.* at 668 ("In the context of higher education, educational diversity is the only non-remedial compelling state interest that courts have sanctioned to date."). See *infra* Part I.B–C. for more on educational diversity specifically.

21. 438 U.S. 265 (1978).

22. See *id.* at 270.

23. See *id.* at 269; see also *id.* at 324 (Brennan, J., concurring); *id.* at 325 (Stevens, J., concurring in part and dissenting in part); Richard D. Kahlenberg, *Class-Based Affirmative Action*, 84 CALIF. L. REV. 1037, 1045 (1996) (referencing "the Court's fractured four-one-four decision in *Bakke*").

24. *Bakke*, 438 U.S. at 319–20 (majority opinion).

25. See *id.* at 315–17.

26. See *id.* at 321 (app.).

interests,²⁷ “[p]ursuit of educational diversity was . . . singled out as a worthy state interest.”²⁸

The vast majority of institutions of higher learning quickly followed Justice Powell’s lead, modeling their own affirmative action policies after the Harvard Plan identified in *Bakke*.²⁹ Yet, twenty years after *Bakke*, appellate courts were still debating the merits of educational diversity as a compelling state interest sufficient to bolster affirmative action.³⁰ One circuit court agreed that educational diversity was a compelling state interest that justified the use of race in admissions.³¹ Another determined both that Justice Powell’s opinion in *Bakke* was not binding precedent and that educational diversity was not a compelling state interest.³² This circuit split set the stage for another Supreme Court affirmative action showdown.³³

A half-century after *Bakke* was decided, the Supreme Court again took up the issue of affirmative action in higher education, granting certiorari in the twin University of Michigan cases *Grutter v. Bollinger*³⁴ and *Gratz v. Bollinger*.³⁵ In those cases, as in *Bakke*, unsuccessful white applicants (to the law school and undergraduate College of Literature, Science, and the Arts, respectively) “complained that including race as a factor in admissions discriminated against them in violation of the Constitution and other anti-discrimination laws.”³⁶ After successful interventions by student and community defendants in both cases,³⁷ a lengthy trial in the law school case,³⁸ a Sixth Circuit summary judgment decision, and a Sixth Circuit rehearing en banc,³⁹ the cases finally reached the Supreme Court. In upholding the law school’s policy in *Grutter*, the Supreme Court confirmed

27. *Id.* at 310.

28. Deo, *supra* note 8, at 669 (citing *Bakke*, 438 U.S. at 316).

29. For instance, the Sixth Circuit opinion in *Grutter* states, “[d]rafted to comply with *Bakke*, the [University of Michigan] Law School’s consideration of race and ethnicity does not use quotas and closely tracks the Harvard plan.” *Grutter v. Bollinger*, 288 F.3d 732, 746 (6th Cir. 2002) (en banc) (quoting law school’s policy), *aff’d*, 539 U.S. 306 (2003).

30. *Smith v. Univ. of Wash.*, 233 F.3d 1188, 1200–01 (9th Cir. 2000); *Hopwood v. Texas*, 78 F.3d 932, 941–44, 948 (5th Cir. 1996).

31. *Smith*, 233 F.3d at 1200–01.

32. *Hopwood*, 78 F.3d at 941–44, 948.

33. See Deo, *supra* note 3, at 68 (“[F]ollowing a circuit split, the Court granted certiorari in *Grutter* in order to give a clear answer to the question of whether institutions of higher learning could rely on affirmative action to improve or maintain student body diversity.” (citing *Grutter*, 539 U.S. at 322 (“We granted certiorari, 537 U.S. 1043 (2002), to resolve the disagreement among the Courts of Appeals on a question of national importance: Whether diversity is a compelling interest that can justify the narrowly tailored use of race in selecting applicants for admission to public universities.”))).

34. 539 U.S. 306 (2003).

35. 539 U.S. 244 (2003).

36. See Deo, *supra* note 8, at 670 (citing *Grutter*, 539 U.S. at 317).

37. See Deo, *supra* note 3, at 70 n.35 (“Student and community intervenors joined the lawsuit to defend affirmative action on grounds of equality and justice rather than rely on the Law School’s more traditional defense of diversity as a compelling state interest.”).

38. See Rachel F. Moran, *The Heirs of Brown: The Story of Grutter v. Bollinger*, in *RACE LAW STORIES* 451, 460–73 (Rachel F. Moran & Devon Wayne Carbado eds., 2008).

39. *Id.* at 473–78.

that educational diversity was a compelling state interest justifying the use of race in admissions decisions in higher education.⁴⁰ The Court also held that the law school's affirmative action policy was sufficiently narrowly tailored to withstand strict scrutiny while the undergraduate policy was not.⁴¹

In the past two years, the Supreme Court has issued two additional opinions relating to affirmative action in higher education. In *Fisher v. University of Texas at Austin*,⁴² decided in 2013, the Court determined that institutions of higher education in Texas could rely on affirmative action to achieve greater educational diversity, along with an existing state program guaranteeing admission to state universities to the top 8–10 percent of students from each high school graduating class.⁴³ Although the “Top 10% Program,” as it became known, “itself produced some student diversity by drawing from highly segregated high schools around the state,”⁴⁴ the Court determined that the university could also assert its interest in educational diversity through direct affirmative action measures, so long as its policy was narrowly tailored to achieve that goal.⁴⁵

Just last year, the Court ruled on *Schuetz v. Coalition to Defend Affirmative Action*,⁴⁶ a challenge to Michigan's Proposal 2 (“Prop. 2”).⁴⁷ Prop. 2 is a Michigan Constitutional Amendment banning public officials from using race as a factor in higher education admissions, as well as in decisions involving hiring and contracting.⁴⁸ When Prop. 2 appeared on the state ballot in 2006, Michigan became the newest battleground to debate affirmative action policy through the popular initiative process.⁴⁹ By then, the modern affirmative action debate had spread beyond the judicial branch.⁵⁰

40. See *Grutter*, 539 U.S. at 325 (“[S]tudent body diversity is a compelling state interest that can justify using race in university admissions.”).

41. See *Gratz*, 539 U.S. at 275 (“[B]ecause the University's use of race in its current freshman admissions policy is not narrowly tailored to achieve respondents' asserted compelling interest in diversity, the admissions policy violates the Equal Protection Clause of the Fourteenth Amendment.”); see also *Grutter*, 539 U.S. at 343 (finding that the “Law School's narrowly tailored use of race in admissions decisions to further a compelling interest in obtaining the educational benefits that flow from a diverse student body” is not prohibited by the Equal Protection Clause).

42. 133 S. Ct. 2411 (2013).

43. *Id.* at 2414.

44. Deo, *supra* note 8, at 672.

45. See *Fisher*, 133 S. Ct. at 2421 (remanding the case to the Fifth Circuit with a directive that “the Court of Appeals must assess whether the University has offered sufficient evidence that would prove that its admissions program is narrowly tailored to obtain the educational benefits of diversity”).

46. 134 S. Ct. 1623 (2014).

47. *Id.* at 1629; see also MICH. CONST. art. I, § 26.

48. See *Schuetz*, 134 S. Ct. at 1629 (“Under the terms of the amendment, race-based preferences cannot be part of the admissions process for state universities.”).

49. For a thorough history of the process by which groups and individuals sought to ban affirmative action through the popular initiative process, see generally Meera E. Deo, *Ebbs and Flows: The Courts in Racial Context*, 8 RUTGERS RACE & L. REV. 167 (2007).

50. *Id.*

Using the popular initiative process as a tool against affirmative action originated in California.⁵¹ Affirmative action became a topic of public conversation in earnest in 1996 when the California Civil Rights Initiative was placed on the state ballot as Proposition 209 ("Prop. 209").⁵² Prop. 209 proposed a state constitutional amendment, which was styled as a "Prohibition Against Discrimination or Preferential Treatment by State and Other Public Entities," and it sought to eliminate race-conscious decision making in public education, employment, and contracting decisions.⁵³ In November 1996, Prop. 209 became law with the support of the majority of California voters.⁵⁴ The constitutional amendment survived legal challenges and went into effect, inspiring a number of other states to consider similar measures.⁵⁵

When Michigan voters passed Prop. 2, supporters of affirmative action quickly responded in court, "alleging that the new law amounts to political restructuring in violation of the Fourteenth Amendment Equal Protection Clause."⁵⁶ In April 2014, in its majority opinion in *Schuette*, the Court did not directly address the benefits of educational diversity, but it held that Michigan's voter initiative passed constitutional muster and could go into effect.⁵⁷ Thus, the *Schuette* Court left educational diversity untouched, allowing it to remain good law, while chipping away at affirmative action more broadly.

C. The Current State of Educational Diversity

The compelling state interest of educational diversity gives institutions of higher learning the freedom to determine how to optimize student learning through their admissions processes.⁵⁸ Many institutions are forthright about their belief that attracting, admitting, and enrolling a diverse student body

51. See *id.* at 179.

52. Don Lattin, *Clergy Denounces Prop. 209: CCRI Called 'Deep Lie' Created to Kill Affirmative Action*, S.F. CHRON., Oct. 22, 1996, at A15.

53. See, e.g., CAL. BALLOT PAMPHLET: GEN. ELEC. NOV. 5, 1996, PROPOSITION 209: PROHIBITION AGAINST DISCRIMINATION OR PREFERENTIAL TREATMENT BY STATE AND OTHER PUBLIC ENTITIES 30–33, available at http://librarysource.uchastings.edu/ballot_pdf/1996g.pdf (outlining Prop. 209 which was later enacted as CAL. CONST. art. I, § 31).

54. Jerome Karabel & Lawrence Wallack, *Proponents of Prop. 209 Misled California Voters*, CHRISTIAN SCI. MONITOR, Dec. 5, 1996, at 19. For more on the history and process surrounding passage of Prop. 209, see Deo, *supra* note 49, at 179–80.

55. Deo, *supra* note 49, at 180.

56. Deo, *supra* note 8, at 666.

57. See *Schuette v. Coal. to Defend Affirmative Action*, 134 S. Ct. 1623, 1630 (2013) ("The question here concerns not the permissibility of race-conscious admissions policies under the Constitution but whether, and in what manner, voters in the States may choose to prohibit the consideration of racial preferences in governmental decisions, in particular with respect to school admissions.").

58. See *Fisher v. Univ. of Tex. at Austin*, 133 S. Ct. 2411, 2419 (2013) (explaining that "some, but not complete, judicial deference is proper" because institutions of higher education deserve academic freedom with regard to the particular goals and mission of the institution).

improves learning outcomes.⁵⁹ The Supreme Court has asserted that educational diversity “promotes ‘cross-racial understanding,’ helps to break down racial stereotypes, and ‘enables [students] to better understand persons of different races.’”⁶⁰ In addition, the Court has asserted that increased campus diversity results in better in-class learning because “classroom discussion is livelier, more spirited, and simply more enlightening and interesting” when students have “the greatest possible variety of backgrounds.”⁶¹

In spite of these assertions, empirical legal scholarship has shown that structural diversity—for instance, having sufficient numbers of students of color on campus—is only a necessary condition for the anticipated improved learning outcomes of educational diversity.⁶² Structural diversity is not sufficient to create the enhanced classroom experience that courts and faculty members expect will result. While meaningful numbers of students of color on campus are *necessary* for cross-racial conversations to occur in the classroom or elsewhere on campus, the numbers alone are not *sufficient* to produce the expected benefits of educational diversity.⁶³ To achieve those results, facilitators must foster positive cross-racial conversations and experiences between those diverse students, “where individuals interact as equals in a mutually respectful environment;” this will maximize the benefits of structural diversity in the classroom.⁶⁴ Given these findings, one would expect greater emphasis on educational diversity in courts and in classrooms, as administrators, educators, and policy makers strive to make the most out of whatever structural diversity exists on campus.

Yet, educational diversity as a legal justification for affirmative action is itself in a precarious position.⁶⁵ Even in *Grutter*, which confirmed that universities could act on their commitment to educational diversity through affirmative action policies, the Court mentioned its expectation “that 25 years from now, the use of racial preferences will no longer be necessary to further the interest [of diversity].”⁶⁶ The Court in *Fisher* more recently made clear that “strict scrutiny imposes on the university the ultimate

59. Even the Law School Admissions Council reports that “[l]aw schools want diverse, interesting classes, representative of a variety of backgrounds.” See *How Law Schools Select Applicants*, LSAC, <http://www.lsac.org/jd/choosing-a-law-school/canadian/law-schools-select-applicants> (last visited Apr. 23, 2015).

60. *Grutter v. Bollinger*, 539 U.S. 306, 330 (2003) (alteration in original).

61. *Id.*

62. Deo, *supra* note 3, at 77 (citing studies finding that “[t]o reap the maximum benefits from educational diversity, students from diverse backgrounds must engage in ‘meaningful interaction’ that is both frequent and of a high quality”).

63. *Id.* at 73 (“Simply admitting students of color in raw numbers (even numbers sufficient to constitute a welcoming atmosphere for those students) is no guarantee that the interactions and classroom conversations the *Grutter* Court anticipated will actually take place.”).

64. *Id.* at 63, 83–84, 84 n.133 (citing Gregory M. Herek, *Myths About Sexual Orientation: A Lawyer’s Guide to Social Science Research*, 1 LAW & SEXUALITY REV. 133, 171 (1991)).

65. See, e.g., Deo, *supra* note 8, at 662 (“Educational diversity [is] resting on shaky ground.”).

66. See *Grutter*, 539 U.S. at 343.

burden of demonstrating, before turning to racial classifications, that available, workable race-neutral alternatives do not suffice.”⁶⁷ Plaintiffs have not been complacent as recent affirmative action challenges have directly attacked educational diversity as a compelling state interest. At least one Supreme Court Justice has made clear that he personally does not believe that educational diversity should continue to be a compelling interest.⁶⁸ As I have written in an earlier article, “with educational diversity resting on shaky ground, we have reached the hour of both bolstering educational diversity and considering viable alternatives.”⁶⁹ This Article draws from empirical data to bolster diversity, contributing the faculty perspective to existing theoretical and empirical investigations into this last remaining non-remedial compelling state interest.

II. EMPIRICAL SCHOLARSHIP ON EDUCATIONAL DIVERSITY

A. Student Attitudes

The past decade has seen an explosion of research on student diversity, especially scholarship utilizing empirical data drawn from students themselves.⁷⁰ Understanding the student perspective on diversity is a critical part of the puzzle. Yet, these pieces alone cannot complete the picture. This Article builds on existing scholarship investigating how students in higher education, and law students in particular, navigate diversity through an examination of law faculty perspectives on diversity.

Research with law students makes clear that “students from all race/ethnic backgrounds not only appreciate diversity, but [also] would prefer greater diversity on campus in order to improve their learning of legal concepts and benefit them in their future careers.”⁷¹ In fact, one study of the law student experience drawing on surveys from over 8000 entering law students across the country discovered that “a whopping 88% of all law student respondents in the national sample support diversity.”⁷²

In terms of the reasons that students give for supporting diversity, the “vast majority” of students in the national sample believe that “diversity enhances their learning environment.”⁷³ A case study of University of Michigan Law School students reveals that those students specifically prefer diversity on campus.⁷⁴ The study cites a number of educational

67. See *Fisher v. Univ. of Tex. at Austin*, 133 S. Ct. 2411, 2420 (2013).

68. *Id.* at 2424–29 (Thomas, J., concurring).

69. Deo, *supra* note 8, at 662.

70. See, e.g., Patricia Gurin et al., *Diversity and Higher Education: Theory and Impact on Educational Outcomes*, 72 HARV. EDUC. REV. 330 (2002); Meera E. Deo et al., *Struggles & Support: Diversity in U.S. Law Schools*, 23 NAT’L BLACK L.J. 71, 77 (2010); Deo, *supra* note 3, at 66.

71. Deo, *supra* note 8, at 690.

72. Deo et al., *supra* note 70, at 81–82.

73. *Id.* at 89.

74. See Walter R. Allen & Daniel Solórzano, *Affirmative Action, Educational Equity, and Campus Racial Climate: A Case Study of the University of Michigan Law School*, 12 BERKELEY LA RAZA L.J. 237, 298 (2001) (“The majority of respondents (68 percent) stated a

benefits of diversity, including that “a) greater structural diversity leads to increased classroom diversity and improved learning; b) classroom diversity results in open minds and engaging classroom conversations; and c) more structural diversity leads to greater participation and less tokenism.”⁷⁵

In addition to appreciating diversity, students also appreciate and seek out *diversity discussions* in class.⁷⁶ Diversity discussions are “classroom conversations regarding race, gender, and/or sexual orientation,” brought up by either faculty or students that are often used to augment lectures on black letter law.⁷⁷ Interestingly, it is not only students of color who appreciate diversity discussions in class; rather:

students from all racial and ethnic backgrounds (89% of White respondents, 82% of Latino respondents, 78% of API respondents, 77% of Other respondents, 75% of Native American respondents, and 73% of Black respondents) agree that they themselves are supportive “when faculty include discussions of race, gender, or sexual orientation in the classroom.”⁷⁸

Many of these students appreciate how including “social context may help [explain] complex issues of law by making them come alive through personal experiences.”⁷⁹ In fact, a majority of law students from all racial/ethnic backgrounds “not only appreciate diversity discussions, but also wish that they were included more often as a standard part of the first-year curriculum.”⁸⁰ While students of all backgrounds appreciate and benefit from diversity discussions, these conversations may be especially important for students of color, who are less likely to be comfortable sharing their personal experiences in the classroom and often uncomfortable drawing from past experience to elucidate particular points of law.⁸¹ In part, this hesitation comes from students of color being underrepresented on

preference for studying with students of diverse racial backgrounds when preparing for examinations.”).

75. Deo, *supra* note 3, at 97.

76. *Id.* at 95.

77. *Id.* Research in this area indicates that law faculty of color and female faculty may be more adept at facilitating diversity discussions in class than their white male colleagues, thus serving the students both what they desire and an arguably more nuanced understanding of the law. See Meera E. Deo et al., *Paint by Number? How the Race and Gender of Law School Faculty Affect the First-Year Curriculum*, 29 CHICANO-LATINO L. REV. 1, 17 (2010) (“Data reveal a pattern based on race and gender such that female faculty and faculty of color are more likely to engage in these discussions, while white male faculty not only are more likely to disregard the racial/gender context of the law but may even be insensitive to diversity issues, contributing to a more challenging environment for some students of color and female students.”).

78. Deo, *supra* note 3, at 95.

79. *Id.*

80. Deo et al., *supra* note 77, at 31.

81. See, e.g., Meera E. Deo, *Separate, Unequal, and Seeking Support*, 28 HARV. J. ON RACIAL & ETHNIC JUST. 9, 19 (2012) (discussing a lack of engagements from students of color who feel marginalized in class).

campus, which feeds into alienation and marginalization on predominantly white campuses.⁸²

In addition to classroom benefits of diversity, many law students “also see significant benefits to their future careers that are based on the benefits of educational diversity in law school.”⁸³ In large part, “diversity in higher education prepares students for interaction in an increasingly globalized workplace.”⁸⁴

Student perspectives are clearly relevant here as empirical research with students can reveal benefits or drawbacks of student diversity, as well as ways to elicit more meaningful cross-racial interaction in the classroom.⁸⁵ In addition to the law student experience, law faculty perspectives also offer unique insights into educational diversity, both in terms of educational benefits and professional benefits, as faculty generally lead classroom conversations and are in tune with whatever level of structural diversity they have in each class.

B. Faculty Preferences

While “[i]t may be no surprise that students from all racial/ethnic backgrounds support diversity and prefer greater levels of diversity on campus,” law faculty perspectives of educational diversity have never been systematically examined.⁸⁶ Law students have offered their perspectives on law faculty, appreciating the benefits they receive from faculty mentorship⁸⁷ and the ways in which nontraditional faculty include diversity discussions in class.⁸⁸ Yet, until now, researchers have never conducted a parallel investigation into law faculty perspectives on law student diversity.

This omission is largely due to the fact that no formal, empirical, comprehensive analysis has been conducted with law faculty at all points of the professional spectrum, at geographically diverse schools, and including racial, ethnic, and gender diversity. A few recent studies have begun to fill this surprising hole in the literature by illuminating some aspects of the law faculty experience. For instance, *Presumed Incompetent* is a recently published anthology of scholars from various disciplines revealing a pattern of challenges facing women of color academics.⁸⁹ Many of these barriers stem from the external assumption of incompetence facing these nontraditional (i.e., nonwhite and non-male) academics even before they begin formal teaching, scholarship, or service on campus.⁹⁰

82. Deo, *supra* note 3, at 77.

83. Deo, *supra* note 8, at 688.

84. *Id.* at 689.

85. Allen & Solórzano, *supra* note 74; Deo, *supra* note 3, at 73; Gurin et al., *supra* note 70, at 333.

86. Deo, *supra* note 8, at 687.

87. Deo, *supra* note 70, at 86–88.

88. Deo, *supra* note 3, at 95; Deo et al., *supra* note 77, at 17.

89. PRESUMED INCOMPETENT: THE INTERACTIONS OF RACE AND CLASS FOR WOMEN IN ACADEMIA (Gabriella Gutiérrez y Muhs et al. eds., 2012).

90. *Id.*

Another ongoing study utilizes empirical data gathered from tenured law faculty to investigate the tenure process specifically.⁹¹ That study discovered that “despite significant progress toward more diversity, women and scholars of color face continued difficulties” in the law school setting.⁹² In fact, when evaluating satisfaction with the tenure process, the study discovered a “much higher percentage of female professors of color reporting the tenure process as unfair (35%) as compared to white males (12%).”⁹³ The authors also reference studies showing ongoing “disparities in terms of pay, tenure denials, and employment at the most elite law schools, in addition to double standards in assessing identical credentials.”⁹⁴

New scholarship on the law faculty experience also draws from the same dataset explored in this Article to show how law faculty experiences differ by race and gender.⁹⁵ For instance, while faculty overall describe interactions with colleagues as cordial, many women of color law faculty members see beyond a mask of civility to underlying hostility from many of their colleagues; this in turn breeds distrust.⁹⁶ Both white women and women of color complain of silencing during faculty meetings, where male colleagues often hold court.⁹⁷ Even more common is the phenomenon of “mansplaining,” where female faculty contributions are literally unheard unless and until a male colleague repeats and takes credit for her ideas.⁹⁸ Direct student confrontations against professors in the classroom—primarily verbal challenges to authority—are aimed almost exclusively at female professors,⁹⁹ as are negative teaching evaluations commenting on personal style (e.g., dress and hair) over substance (e.g., teaching effectiveness).¹⁰⁰ There are also documented barriers to entry blocking nontraditional candidates from entering legal academia, as well as challenges preventing women of color from assuming formal administrative leadership positions.¹⁰¹ While implicit bias may account for many of these negative experiences and outcomes, overt discrimination also contributes to the many challenges facing nontraditional law faculty members.¹⁰²

As outlined above, there is a growing body of scholarship investigating the law faculty experience, much of which comes from the same dataset

91. Katherine Barnes & Elizabeth Mertz, *Is It Fair? Law Professors' Perceptions of Tenure*, 61 J. LEGAL EDUC. 511, 512 (2012).

92. *Id.*

93. *Id.* at 511.

94. *Id.* at 512.

95. For details on DLA, see *infra* Part II.C.

96. See Meera E. Deo, *The Ugly Truth About Legal Academia*, 80 BROOK. L. REV. (forthcoming 2015).

97. See *id.*

98. *Id.* (describing mansplaining).

99. See *id.*

100. See Meera E. Deo, *A Better Tenure Battle*, 28 COLUM. J. GENDER & L. (forthcoming 2015).

101. See Meera E. Deo, *Trajectory of a Law Professor*, MICH. J. RACE & L. (forthcoming 2015).

102. See Deo, *supra* note 96; Deo, *supra* note 100.

presented in this Article.¹⁰³ Yet, none of the existing literature touches on law faculty perspectives regarding student diversity. This Article incorporates the law faculty experience into the broader educational diversity conversation. Doing so provides much-needed empirical support for diversity as a compelling state interest from the very individuals in the best position to offer it. Since law faculty are the ones leading classrooms, they are best able to determine the effectiveness of including diversity discussions in class, express the benefits of engaging with diverse students who draw from personal experience to illuminate the law, and facilitate respectful conversations among students from different backgrounds so that they can truly learn from the diversity in their midst.

C. *The Diversity in Legal Academia Project*

The empirical data presented in this Article are drawn from the Diversity in Legal Academia (DLA) project. DLA provides empirical data on a range of issues related to the personal and professional lives of law faculty members. As the principal investigator of the DLA project, I not only personally conducted all ninety-three one-on-one interviews with legal academics, but I also maintain responsibility for all ongoing elements of the project from data collection to coding, analysis, and dissemination.¹⁰⁴ DLA includes survey and interview data from ninety-three legal academics around the United States who are diverse according to a number of different *domains*, including: race/ethnicity, gender, tenure status, leadership, location/region of school, and institutional selectivity. DLA participants are tenured and tenure-track faculty employed at AALS-member¹⁰⁵ and ABA-accredited law schools.¹⁰⁶ The *core sample* of DLA is women of color (nonwhite women), including sixty-three female law faculty members who are Black, Latina, Asian American, Native American, Middle Eastern, and multiracial.¹⁰⁷ The study also incorporates perspectives from a *comparative*

103. See, e.g., PRESUMED INCOMPETENT, *supra* note 89; Barnes & Mertz, *supra* note 91. For additional articles drawing from DLA data, see Deo, *supra* note 101; Deo, *supra* note 100.

104. For a primer on these aspects of mixed-methods research design and operationalization, see JOHN W. CRESWELL & VICKI L. PLANO CLARK, DESIGNING AND CONDUCTING MIXED METHODS RESEARCH 1 (2007); ABBAS TASHAKKORI & CHARLES TEDDLIE, MIXED METHODOLOGY: COMBINING QUALITATIVE AND QUANTITATIVE APPROACHES 5 (1998).

105. The Association of American Law Schools (AALS) is a 501(c)(3) nonprofit association currently consisting of 178 member law schools, including “most of the nation’s law students [which] produce the majority of the country’s lawyers and judges, as well as many of its lawmakers.” *About*, ASS’N OF AM. LAW SCH., <http://www.aals.org/about/> (last visited Apr. 23, 2015).

106. Clinical, legal writing, and library faculty are excluded from participation in DLA, consistent with other empirical and theoretical research on law faculty, because their experiences tend to differ substantially from those of other law faculty members. For more on why these faculty members are not included and for a list of other scholarship differentiating between faculty by status, see Meera E. Deo, *Looking Forward to Diversity in Legal Academia*, 29 BERKELEY J. GENDER L. & JUST. 352, 377 n.167 (2014).

107. Race is a fluid concept, which often defies easy characterization or categorization. See, e.g., Ian F. Haney López, *The Social Construction of Race: Some Observations on*

sample of thirty white women, white men, and men of color. Table 1 provides details on DLA participants by race/ethnicity and gender.

Table 1: DLA Participants, by Race and Gender, DLA 2013 (N=93)

		Male	Female	Total
Black	N	4	21	25
	%	16.00%	84.00%	100.00%
Asian/Pacific Islander	N	3	15	18
	%	16.67%	83.33%	100.00%
Latino/Latina	N	2	13	15
	%	13.33%	86.67%	100.00%
Native American	N	1	5	6
	%	16.67%	83.33%	100.00%
Middle Eastern	N	1	2	3
	%	33.33%	66.67%	100.00%
Multiracial	N	1	7	8
	%	12.50%	87.50%	100.00%
White	N	7	11	18
	%	38.89%	61.11%	100.00%
Total	N	19	63	93

Data collection for DLA followed a *target sample* approach. Target sampling is a variation of *snowball sampling*, a common methodological tool pioneered by statisticians and also used by scholars in the social sciences.¹⁰⁸ Target sampling provides greater structure and representativeness than a standard snowball sample through constant monitoring and direct intervention of the dataset throughout the data collection period.¹⁰⁹

Illusion, Fabrication, and Choice, 29 HARV. C.R.-C.L. L. REV. 1, 7 (1994) (drawing on research from various fields to “repudiate the idea that race is a fixed essence and instead locate[s] races within the cartography of other social constructions”). DLA uses self-identification of participants in both the survey and interview. The terms “African American” and “Black” are used interchangeably throughout the Article to refer to those who self-identified using those terms. Participants who identified as “API,” “Asian,” “Asian American,” or within one of the pan-ethnic Asian-American identities are identified as “Asian American,” while those who self-identified as “Latino” or “Hispanic” are referred to as “Latino.” Those who identified only as “white” are identified as such in the Article. Multiracial participants are those who self-identified as having two or more racial/ethnic backgrounds. These racial/ethnic categories were chosen because they are the ones used by AALS in their statistics.

108. See, e.g., Leo A. Goodman, *Snowball Sampling*, 32 ANNALS MATHEMATICAL STAT. 148, 148 (1961).

109. See Douglas D. Heckathorn, *Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations*, 44 SOC. PROBS. 174, 175 (1997) (explaining the potential biases affecting snowball sampling); John K. Watters & Patrick Biernacki, *Targeted Sampling: Options for the Study of Hidden Populations*, 36 SOC. PROBS. 416, 420 (1989).

Target sampling specifically begins with a *seed group* of participants that is carefully selected to be broadly representative of the larger population; in this case, the population is law faculty members who are diverse with regard to each of the domains mentioned above.¹¹⁰ In the DLA study, seed participants completed a survey asking broad demographic, attitudinal, and experiential questions before continuing on to personal one-on-one interviews with me, the principal investigator of DLA. The penultimate question on the survey followed standard protocol for snowball and target sampling by asking seed group participants to nominate others they thought might be interested in joining the study.¹¹¹ By carefully reviewing existing representation of the sample, I then selected from among those nominated to maintain representation in the final sample along all of the domains mentioned.¹¹² In this way, target sampling is “an ongoing and interactive process in which data are constantly analyzed and used to adjust the recruitment and sampling techniques.”¹¹³ Any oversampling of a particular group (e.g., three Black female professors from the Midwest) can easily be corrected in the final sample by drawing more from areas that are underrepresented (e.g., one additional Black female professor from the East Coast) as the study continues.

Thus, just as a snowball grows by rolling through snow that sticks to and enlarges the original collection of snow, so the sample grows by adding research subjects drawn from names suggested by prior participants. The target approach allows for careful calibration of the sample to ensure that it remains representative throughout the process and especially in final form.¹¹⁴

Target sampling is an especially useful and effective data collection method when the target population is hidden or vulnerable, making these individuals otherwise unlikely or unable to participate in formal empirical projects.¹¹⁵ The methodological approach utilized for this study focused on women of color law faculty members because this group is also surprisingly hard to identify¹¹⁶ and disinclined to participate in formal empirical studies emphasizing their nontraditional status in legal academia.¹¹⁷

(defining targeted sampling as “a purposeful, systematic method by which controlled lists of specified populations within geographical districts are developed and detailed plans are designed to recruit adequate numbers of cases within each of the targets”).

110. Watters & Biernacki, *supra* note 109.

111. *Id.* at 420.

112. Again, the tracked domains are race/ethnicity, gender, tenure status, leadership, location/region of school, and institutional selectivity.

113. *Id.* at 421.

114. *Id.* at 420.

115. *Id.* (explaining that snowball sampling is often used in studies of homeless youth, those infected with HIV, and with other populations that are difficult to identify or not inclined to participate in formal empirical studies).

116. To start, there is little reliable data on law faculty. The data we have on law faculty members, disaggregated by race/ethnicity and gender, is from 2008 to 2009; if I used that data for the DLA study, I would essentially exclude most assistant professors. AALS also publishes a directory of law professors, and members can opt in to a “Minority Law Professor” listing, though faculty are not identified by race/ethnicity or gender in that listing.

I personally conducted each interview for the DLA study, either in person or by telephone. Each interview was digitally audiotaped and transcribed before coding and analysis began.¹¹⁸ The interview protocol includes experiential questions (i.e., regarding interactions with colleagues and students), attitudinal questions (i.e., regarding perspectives on diversity and institutional values), and more personal demographic questions (i.e., regarding work/life balance and the participant's professional trajectory).

A comprehensive codebook guides data coding and analysis.¹¹⁹ Following emerging theme analysis, I began identifying patterns while data collection was ongoing.¹²⁰ Codes were then developed to more systematically identify patterns throughout the data.¹²¹ For instance, most mothers in the DLA sample discussed challenges balancing professional and personal obligations, even when partners were very involved in childcare. Thus, I began coding for childcare specifically. Additional codes were also developed using the interview protocol (the list of actual interview questions) as a guide. For instance, because each participant was asked about work/life balance generally, I also created a coding scheme related to balancing personal and professional obligations.¹²²

Formal qualitative coding and analysis for this project utilizes ATLAS.ti, a qualitative coding software program that facilitates organization of the data.¹²³ Quantitative data were analyzed with Stata and Excel for basic

Visual identification of faculty to determine race/ethnicity (through researcher-identified race/ethnicity based on characterization of an online picture associated with each faculty member's institutional biography) would be highly problematic, as race/ethnicity is such a fluid concept, and it would perhaps be problematic for gender identification as well. See, e.g., Aliya Saperstein & Andrew M. Penner, *Beyond the Looking Glass: Exploring Fluidity in Racial Self-Identification and Interviewer Classification*, 57 SOC. PERSPS. 186 (2014).

117. As noted in a previous article laying the methodological and other foundations for DLA:

The Introduction to *Presumed Incompetent* includes an entire section explaining that many women were uncomfortable publishing their stories, and therefore did not participate by submitting non-anonymous narratives to the anthology. Some of these women were too emotionally exhausted to think critically about their experiences and share them publicly; many others feared professional repercussions of "outing" their negative experiences and the institutions in which those experiences occurred.

See Deo, *supra* note 106, at 380 n.180 (citing PRESUMED INCOMPETENT, *supra* note 89, at 10–14).

118. This is standard practice for mixed-method empirical research. See CRESWELL & CLARK, *supra* note 104, at 1; TASHAKKORI & TEDDLIE, *supra* note 104, at 5.

119. This is standard practice for mixed-method empirical research. See CRESWELL & CLARK, *supra* note 104, at 1; TASHAKKORI & TEDDLIE, *supra* note 104, at 5.

120. See ROBERT M. EMERSON, CONTEMPORARY FIELD RESEARCH PERSPECTIVES AND FORMULATIONS 284 (2d ed. 2001) (explaining the rationale behind using emerging theme analysis).

121. This is standard practice for mixed-method empirical research. See CRESWELL & CLARK, *supra* note 104, at 1; TASHAKKORI & TEDDLIE, *supra* note 104, at 5.

122. This is standard practice for mixed-method empirical research. See CRESWELL & CLARK, *supra* note 104, at 1; TASHAKKORI & TEDDLIE, *supra* note 104, at 5.

123. ATLAS.TI, <http://www.atlasti.com/index.html> (last visited Apr. 23, 2015); see also CHRISTINA SILVER & ANN LEWINS, USING SOFTWARE IN QUALITATIVE RESEARCH (2d ed.

cross-tabulations of the data as well as more advanced analyses not presented in this Article.¹²⁴ Dissemination of preliminary and formal findings has been ongoing through presentations at numerous academic conferences and various invited speaking engagements.¹²⁵

Past publications utilizing DLA data have focused on the ways in which race/ethnicity and gender create or lead to different experiences for various law faculty. For instance, women of color endure significant verbal and physical confrontations in the classroom from students, though these tend to be absent in classrooms taught by white men.¹²⁶ In addition, a set of individual and structural barriers often prevent nontraditional faculty from assuming formal administrative leadership positions in legal academia, though white men rarely face these hurdles.¹²⁷

This Article, unlike others using DLA data, explores law faculty as a group, highlighting their common support for educational diversity. Thus, while the race/ethnicity and gender of participants is included in the findings presented in this Article, there is little variation in terms of preferences based on race/ethnicity and gender.¹²⁸ Law faculty members, both men and women who are white, Black, Latino, Asian American, Middle Eastern, Native American, and multiracial, express a preference for educational diversity and provide surprisingly common rationales and experiences to support their leanings. Thus, the findings presented in the next part are grouped by theme rather than by the race/ethnicity of the participant, highlighting the educational and professional benefits of diversity as expressed by law faculty members from various backgrounds. Names used throughout this Article are pseudonyms to protect the anonymity of DLA participants.¹²⁹ The qualitative data presented are the actual words of the law faculty participants in the DLA project to give voice to their actual experiences, attitudes, and preferences.

2007) (step-by-step guide explaining the functions of the major software programs that are generally used in studies of this kind, including ATLAS.ti).

124. Stata is a statistical software package developed and sold by StataCorp. See STATA, <http://www.stata.com> (last visited Apr. 23, 2015); see also ALAN C. ACOCK, A GENTLE INTRODUCTION TO STATA (4th ed. 2014) (providing an introduction to Stata for researchers working in psychology, social sciences, and other fields that require quantitative analysis).

125. For instance, I have made presentations drawing from DLA data at the AALS Annual Meetings in 2013, 2014, and 2015; as the Neil Gotanda Awardee in Berkeley, California, in 2014; and as an invited speaker at O'Melveny & Myers LLP in Los Angeles, California, in 2014.

126. Deo, *supra* note 96.

127. Deo, *supra* note 101, at 29–30.

128. See *infra* Part III for these empirical findings.

129. This is standard practice for mixed-method empirical research. See CRESWELL & CLARK, *supra* note 104, at 1; TASHAKKORI & TEDDLIE, *supra* note 104, at 5.

III. EMPIRICAL FINDINGS OF LAW FACULTY SUPPORT FOR DIVERSITY

A. Support for Diversity

Quantitative findings from the DLA study are instructive in order to gauge law faculty perspectives on educational diversity. The DLA survey asked participants to rank on a scale of one to five their agreement with a set of statements, providing response options that follow a traditional Likert-scale presentation ranging from Strongly Agree (=5) to Strongly Disagree (=1).¹³⁰ Answers to four questions that are specifically about student diversity and classroom conversations are presented and discussed in this section, disaggregated by race/ethnicity and gender.¹³¹

One question asked law faculty members to rate their level of agreement with the following statement: “My law school student body is as diverse as I expected it to be.” The results, displayed in Table 2, show that many law faculty members are not surprised by the student diversity on their campus—regardless of whether their campus was highly diverse or less so. There are some disparities by race. For instance, 81 percent of Black women, 64 percent of Asian Americans, 58 percent of Latinas, and 60 percent of Native Americans report that their campus is not as diverse as they expected, either disagreeing or strongly disagreeing with the statement. In other words, the majority of women of color in the sample see their school as less diverse than expected. White women have somewhat similar responses, with 45 percent reporting disagreement with the statement. In contrast, only 36 percent of men of color and 27 percent of white men note that existing diversity does not match their expectations. This could mean many things. White men and men of color may be better informed about actual diversity statistics at their schools and so are not surprised by a lack of diversity if they see one. Women as a whole—including women of color and white women—may be more in tune to issues of diversity, making them more likely to consider their expectations as they begin a new position and noting less diversity than expected if that is the case.

130. “Response options commonly referred to as ‘Likert scales’ are used extensively in the social sciences, economics, and other fields to determine respondent attitudes and opinions through their level of agreement with various assertions.” Deo, *supra* note 8, at 686 n.128 (citing Rensis Likert, *A Technique for the Measurement of Attitudes*, 22 ARCHIVES PSYCHOL. 5, 5–6 (1932)).

131. The responses of the core sample of women of color are disaggregated by the race/ethnicity of respondents, including Black, Asian American, Latina, Native American, Middle Eastern, and multiracial women. The comparative samples of white men, white women, and men of color are presented next.

Table 2: *Agreement that Student Body Meets Diversity Expectations, by Race/Ethnicity and Gender, DLA 2013 (N=91)*

		Strongly Agree	Agree	Neither	Disagree	Strongly Disagree	Total
Black Female	N	1	3	0	11	6	21
	%	4.76%	14.29%	0.00%	52.38%	28.57%	100.00%
Asian/Pacific Islander Female	N	1	4	0	8	1	14
	%	7.14%	28.57%	0.00%	57.14%	7.14%	100.00%
Latina	N	0	5	0	5	2	12
	%	0.00%	41.67%	0.00%	41.67%	16.67%	100.00%
Native American Female	N	0	2	0	2	1	5
	%	0.00%	40.00%	0.00%	40.00%	20.00%	100.00%
Middle Eastern Female	N	0	0	0	2	0	2
	%	0.00%	0.00%	0.00%	100.00%	0.00%	100.00%
Multiracial Female	N	0	1	0	3	3	7
	%	0.00%	14.29%	0.00%	42.86%	42.86%	100.00%
White Men	N	0	3	3	2	0	8
	%	0.00%	37.50%	37.50%	25.00%	0.00%	100.00%
White Female	N	0	3	3	5	0	11
	%	0.00%	27.27%	27.27%	45.45%	0.00%	100.00%
Men of Color	N	0	4	3	3	1	11
	%	0.00%	36.36%	27.27%	27.27%	9.09%	100.00%
Total	N	2	25	9	41	14	91
	%	2.20%	27.47%	9.89%	45.05%	15.38%	100.00%

As noted earlier, many law faculty members make efforts to include diversity discussions in class.¹³² The DLA study sought to gauge how faculty members perceive student interest in these conversations about race, gender, and sexual orientation. Earlier research has documented that law students tend to be supportive of diversity discussions, recognizing how they help bring the law to life.¹³³ But no prior study has asked law faculty members whether they thought that students preferred these conversations in the classroom. Specifically, the DLA survey asked participants to rate their level of agreement with the following statement: “My students seem

132. See *supra* Part II.A.

133. Deo et al., *supra* note 77, at 17 (explaining that “students, regardless of race or gender, tend to prefer the approach most often used by female faculty and faculty of color, who actively involve diversity discussions in their law teaching”).

supportive when I include discussions of race, gender, or sexual orientation in the classroom.” Table 3 presents results by race/ethnicity and gender.

Responses to this question indicate differences in how faculty of color and white faculty perceive student interest in diversity discussions. For example, 100 percent of white men and 82 percent of white women believe their students are supportive when they themselves bring diversity discussions into the classroom.¹³⁴ In contrast, only 55 percent of men of color agree that students appreciate when they bring up diversity discussions in class, 27 percent are neutral on the subject, and only 18 percent disagree. Women of color from various racial/ethnic groups are much more likely to disagree with the statement, with roughly 40 percent of Black women, Asian American women, Latinas, and Native American women believing that students are *not* supportive when they bring up issues of race, gender, or sexual orientation in class. Interpreting this finding, however, is tricky. It could mean that white men and white women (and to some extent men of color) are better able to gauge student interest in diversity discussions, recognizing how students appreciate their inclusion in class. More likely, based on biases documented in other scholarship examining the experience of women of color law faculty,¹³⁵ it could mean that students respond differently to inclusion of diversity discussions depending on the race and gender of the faculty member bringing up these sensitive topics. Qualitative DLA data also indicates that students may be more responsive and receptive to white faculty members initiating these conversations than to nonwhite faculty members and especially women of color who bring them up.¹³⁶

Table 3: Perceptions of Student Support for Diversity Discussions, by Race/Ethnicity and Gender, DLA 2013 (N=91)

		Strongly Agree	Agree	Neither	Disagree	Strongly Disagree	Total
Black	N	1	11	0	9	0	21
	%	4.76%	52.38%	0.00%	42.86%	0.00%	100.00%
Asian/Pacific Islander	N	2	6	0	6	0	14
	%	14.29%	42.86%	0.00%	42.86%	0.00%	100.00%
Latino/Latina	N	2	5	0	4	1	12
	%	16.67%	41.67%	0.00%	33.33%	8.33%	100.00%
Native American	N	0	3	0	2	0	5
	%	0.00%	60.00%	0.00%	40.00%	0.00%	100.00%

134. Interestingly, only one white participant in the DLA sample (a woman) was “neutral” on this question, with neutrality suggesting that the law faculty member may not bring up diversity discussions herself at all.

135. Deo, *supra* note 96; *see also* Deo, *supra* note 100.

136. *See infra* at Part III.B.

Middle Eastern	N	0	0	0	2	0	2
	%	0.00%	0.00%	0.00%	100.00%	0.00%	100.00%
Multiracial	N	0	3	0	3	1	7
	%	0.00%	42.86%	0.00%	42.86%	14.29%	100.00%
DLA Women	N	1	7	0	0	0	8
	%	12.50%	87.50%	0.00%	0.00%	0.00%	100.00%
DLA White Females	N	5	4	1	1	0	11
	%	45.45%	36.36%	9.09%	9.09%	0.00%	100.00%
DLA Men of Color	N	2	4	3	1	1	11
	%	18.18%	36.36%	27.27%	9.09%	9.09%	100.00%
Total	N	13	43	4	28	3	91
	%	14.29%	47.25%	4.40%	30.77%	3.30%	100.00%

In addition to asking participants about their perceptions of student support for diversity discussions, the DLA survey also asked participants to rate their level of agreement with the following statement about their colleagues: “Most of my fellow faculty are supportive when faculty include discussions of race, gender, or sexual orientation in the classroom.” The findings, presented in Table 4, suggest that the overwhelming majority of whites (100 percent of white men and 73 percent of white women) believe that their colleagues support faculty inclusion of diversity discussions in class, just as they believe their students do. In contrast, smaller percentages of faculty of color agree, with roughly equivalent statistics for men and women. Roughly 40 percent of Black and Asian American women, 55 percent of men of color, and roughly 60 percent of Latinas and Native American women feel that their colleagues support their inclusion of these conversations in class. This likely says more about challenging faculty interactions than about diversity discussions themselves, as women of color faculty tend to have fraught relationships with colleagues, most of whom are white and often male.¹³⁷

Table 4: Perceptions of Faculty Support for Diversity Discussions, by Race/Ethnicity and Gender, DLA 2013 (N=91)

		Strongly Agree	Agree	Neither	Disagree	Strongly Disagree	Total
Black	N	2	6	0	13	0	21
	%	9.52%	28.57%	0.00%	61.90%	0.00%	100.00%
Asian/Pacific Islander	N	3	3	0	7	1	14
	%	21.43%	21.43%	0.00%	50.00%	7.14%	100.00%

137. For more on challenging faculty interactions, see Deo, *supra* note 96, at 22–45.

Latino/Latina	N	2	5	0	4	1	12
	%	16.67%	41.67%	0.00%	33.33%	8.33%	100.00%
Native American	N	1	2	0	1	1	5
	%	20.00%	40.00%	0.00%	20.00%	20.00%	100.00%
Middle Eastern	N	1	0	0	1	0	2
	%	50.00%	0.00%	0.00%	50.00%	0.00%	100.00%
Multiracial	N	1	4	0	2	0	7
	%	14.29%	57.14%	0.00%	28.57%	0.00%	100.00%
DLA Women	N	1	7	0	0	0	8
	%	12.50%	87.50%	0.00%	0.00%	0.00%	100.00%
DLA White Females	N	5	3	0	2	1	11
	%	45.45%	27.27%	0.00%	18.18%	9.09%	100.00%
DLA Men of Color	N	4	2	1	3	1	11
	%	36.36%	18.18%	9.09%	27.27%	9.09%	100.00%
Total	N	20	32	1	33	5	91
	%	21.98%	35.16%	1.10%	36.26%	5.49%	100.00%

All DLA participants also reported on their satisfaction with existing student diversity on campus. Specifically, they responded to a question on the survey asking them to rate their level of agreement with the following statement: “I would prefer that there were more student diversity at my law school.” In spite of various differences by race/ethnicity with regard to other questions about student diversity and inclusion of diversity discussions, there is relative uniformity across race/ethnicity and gender in response to this question. A full 95 percent of Black women agree with the statement, along with 87 percent of Asian American women, 75 percent of Latinas, and 80 percent of Native American women. When we compare women of color to other groups, the vast majority of white men, white women, and men of color also agree. Thus, Table 5 shows that the vast majority of law professors—virtually everyone included in the DLA sample—prefers greater student diversity at their law school. While various interpretations are available for this finding as well, the qualitative data presented in the next two sections perhaps explain it best; the next section covers multiple benefits to diversity in the educational and professional context as well as the challenges associated with a lack of student diversity on campus.

*Table 5: Preference for Greater Student Diversity,
by Race/Ethnicity and Gender, DLA 2013 (N=91)*

		Strongly Agree	Agree	Neither	Disagree	Strongly Disagree	Total
Black	N	15	4	0	1	0	20
	%	75.00%	20.00%	0.00%	5.00%	0.00%	100.00%
Asian/Pacific Islander	N	10	3	0	2	0	15
	%	66.67%	20.00%	0.00%	13.33%	0.00%	100.00%
Latino/Latina	N	7	2	0	3	0	12
	%	58.33%	16.67%	0.00%	25.00%	0.00%	100.00%
Native American	N	3	1	0	1	0	5
	%	60.00%	20.00%	0.00%	20.00%	0.00%	100.00%
Middle Eastern	N	0	0	0	0	2	2
	%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
Multiracial	N	5	2	0	0	0	7
	%	71.43%	28.57%	0.00%	0.00%	0.00%	100.00%
White Men	N	2	5	0	0	0	7
	%	28.57%	71.43%	0.00%	0.00%	0.00%	100.00%
White Female	N	6	4	1	0	0	11
	%	54.55%	36.36%	9.09%	0.00%	0.00%	100.00%
Men of Color	N	4	6	1	0	0	11
	%	36.36%	54.55%	9.09%	0.00%	0.00%	100.00%
Total	N	52	27	2	7	2	90
	%	57.78%	30.00%	2.22%	7.78%	2.22%	100.00%

B. Benefits of Diversity

The quantitative data presented above frame the qualitative data discussed in this section, reporting actual quotations from in-depth interviews and revealing more nuanced perceptions and preferences of individual law faculty members. This section reports on benefits of diversity as experienced by law faculty members, most of whom see their own law schools as having a diverse student body. Three interrelated themes emerge: (1) educational diversity allows for a richer range of perspectives to be included in the classroom, (2) with personal context helping to illuminate black letter law, and (3) providing benefits that will reach into future legal practice.

1. Diverse Perspectives

“I think students learn from one another just as much as they learn from the professor.” — Imani

The most common student diversity pattern that emerges from the DLA law faculty data is that having a diverse student body leads to diverse perspectives being voiced in the classroom. While this may seem obvious, it is not necessarily a foregone conclusion that structural diversity leads to classroom diversity. In fact, one recent study has shown that even at institutions of higher learning where students have meaningful cross-racial interactions on campus generally, these interactions do not necessarily occur in the classroom.¹³⁸ Nevertheless, structural diversity—meaningful numbers of students of color on campus—is a prerequisite to classroom diversity: without diverse students admitted and enrolled, it is impossible for there to be meaningful conversations where students draw on their race/ethnicity, gender, sexual orientation, class, and other identity characteristics to inform classroom conversations and help their peers learn the law.¹³⁹ As Ryan, a Black male faculty member, says, “[Student diversity] makes the classroom conversation more diverse, brings in more perspectives, allows students to learn from others as opposed to learning from people with the same background as themselves, makes for richer interactions, and builds a wonderful community.” A white female law professor named Jordan agrees that “student diversity is a great thing.” For her, “the best thing it adds is just diversity of the dialogues and perspectives in the classroom setting.” Kayla, a Black female law professor, has an especially rich diversity of students in her Property class, which she notes “is a large lecture” class, but nevertheless enjoys incredible discussion drawing from diverse student backgrounds and experiences. Kayla states:

I teach a Property evening class, a large lecture, and I mean every row is just like sprinkled with queer students, students of color, Asian, Indian, Black, you know, it’s just so amazing. The benefit of [diversity] is [that] I think it, for me, helps to bring the different perspectives in the classroom. I know that sounds cliché but it really does.

Like Kayla, most DLA participants are clear that the myriad perspectives that result from student body diversity go beyond race/ethnicity and gender. Imani, a Black woman, provides a representative response, saying, “Well, I mean, I think the perspectives that diverse students bring—and by diversity, [I mean] not just racial and ethnic and gender diversity, but also socioeconomic diversity—and seeing people from different socioeconomic classes and statuses I think really helps to broaden people’s perspective.” She mentions one particular seminar she teaches where “out of sixteen

138. Deo, *supra* note 3, at 63.

139. For a discussion of the ways in which educational diversity may be considered problematic because it puts students of color in a position where they are seeking an education themselves while also being expected to educate their peers, see *id.* See also Deo, *supra* note 8, at 691.

students, eight of them are African American and eight of them are white.” The critical mass of students of color in that class, where their numeric representation mirrors that of their white peers in a way that it rarely does in other classes or even in communities, improves the classroom environment for everyone.¹⁴⁰ Imani notes that as a result of this diversity:

the conversations that we’ve had have just been exceptional and the fact that everyone’s eyes have been opened by hearing about the experiences of those who may not have the same background as them has been tremendous. And I think those benefits, you definitely don’t get those when you don’t have a diverse student body.

While broader diversity and inclusion of perspectives from Latino, Asian American, Native American, and other underrepresented students could only expand perspectives further, Imani sees her diverse classroom as a good start.

Many faculty members also draw on their own backgrounds to help students master the material.¹⁴¹ Yet, as Imani says, “I think students learn from one another just as much as they learn from the professor.” While there is only so much one faculty member can do to bring in various perspectives, Imani has “seen it firsthand where students had not thought of a particular perspective until that perspective was raised by either a student of a diverse background or a faculty member from a diverse background.” By students bringing up these “personal experiences,” classmates learn that sometimes even when “everything on its face seems one way, [] that’s not always the way it is in practice.” April, a Black law professor, notes that there are frequently conversations about race without racial minorities present. “It would be like talking about a pay equity case, a gender equity case, and having no women in the room.” While she thinks that would be obviously problematic, “[i]t happens all the time with respect to race. There are no minorities in the room, so when we look at a case, the only one who can talk about the minority perspective or what might be different, the sensitivity you might need, would be me. And nobody wants to hear me.” As a result, she is confident that “the classroom could be richer if we had more voices.” Since she can only work with the racial/ethnic diversity that exists in the classroom, she suggests students find other ways to diversify their experiences with learning the law outside of class. April continues:

So even in study groups when I counsel students and I talk to them about forming study groups, I suggest to them that they find people who they are not like because if you all come from the same place, you will see the

140. Perhaps conversation would be even more fulfilling with greater racial/ethnic diversity from Latinos, Asian Americans, and others who could contribute beyond the Black/white binary. See generally Juan F. Perea, *The Black/White Binary Paradigm of Race: The “Normal Science” of American Racial Thought*, 85 CAL. L. REV. 1213 (1997).

141. See *supra* Tables 2–3 and accompanying discussion (discussing law faculty members’ perspectives of how students and colleagues view their decision to include diversity discussions in class); see also *infra* Part III.B.2 (discussing additional detail on how having students, rather than faculty members, initiate these discussions is preferable for learning).

problem in the same way, you will begin your analysis from the same point. You have to pick somebody who can see it differently because that challenges your conviction, your belief. It's true in terms of race, class, gender, sexuality. It's true. We all approach things from where we come from, from how life has informed us. It doesn't mean that the diverse voice is more knowledgeable or has a deeper insight or a better ability to judge and determine what's right, what I'm saying is they have a different lens, which can help you challenge your own view as the majority person.

In addition, widespread participation and sharing personal experiences in class breaks down stereotypes between people from different backgrounds who may not have had many prior opportunities for meaningful interaction. White students come to understand that their student of color classmates are just as intelligent, qualified, and capable as they are, while students of color realize that their white classmates may have endured challenges in their own lives too. Matt, a white male law professor, also appreciates the students of color in his classes who draw from personal experience to help their fellow classmates understand a different perspective. Matt says, "I think it's just important for [both] minority and nonminority students to realize the capability of everyone in their cohort and get different perspectives of folks in their cohort." One way he maximizes participation and the sharing of experiences drawing from personal background is simple: "I also teach Criminal Procedure and I ask for a show of hands of students or people they know, so I don't single students out, of people who have been stopped and frisked." The results do not surprise him, though they "may be surprising to some of the students" because the overwhelming majority of students of color raise their hands, as do some whites.

In these ways, white law faculty as well as faculty of color tend to bring up diversity discussions in class, as referenced earlier in this Article.¹⁴² Ellen, a white female, provides some insight into the disparities between how faculty of color and white faculty perceive their efforts at including conversations about race/ethnicity, gender, and sexual orientation in the classroom. She says, "[A]s a white faculty member, I feel that I can more easily engage issues of race than my colleagues of color." Faculty of color may get pushback from students who see them as promoting a race-based agenda or confusing their personal experiences with their professional obligations as law professors.¹⁴³ This helps explain why higher percentages of white faculty than faculty of color believe students appreciate when they bring in diversity discussions.¹⁴⁴ Because Ellen feels she can comfortably bring up these conversations in class without pushback from students, and she sees the value in doing so, noting, "I try to do that [even though] it's

142. See *supra* Tables 2–3 and accompanying discussion (discussing law faculty members' perspectives of how students and colleagues view their decision to include diversity discussions in class).

143. In fact, in anonymous student evaluations, many students accuse faculty of color and female faculty in particular of doing just this. For more on this issue, see Deo, *supra* note 100.

144. See *supra* Table 3 and accompanying discussion.

awkward to do that in a lot of ways,” because she cannot herself put forth multiple theoretical perspectives as persuasively as those who have actually lived them.¹⁴⁵

2. Personal Context

Diversity “opens the doors for a great conversation.” — Ryan

In fact, this opportunity for students to learn from the personal experiences of their classmates is one of the main benefits of educational diversity, as observed by law faculty in the DLA study. Many explain how they can draw on students’ backgrounds to provide personal context that elucidates complex black letter law. Erin, a Native American law professor, says, “I teach Property this semester and I have a Hispanic man in my Property class. And we taught this *State v. Shack*¹⁴⁶ case and it’s about some immigrant workers working on a farm and the right of entry, well it’s going into the right to exclude.” While Elaine could clearly lecture about the substantive area of law without discussing the immigrant experience or the human element of the parties to the case, she notes that including this personal context plays a huge part in classroom learning. She says of the student who spoke up in class that “it was really nice to have him in class because he was willing to talk about the immigrant workers and his family’s background in that.” In other words, he made it personal, which brought the law to life for everyone else. She continues: “I think he gave his classmates a lot more insight than they otherwise would have because they wouldn’t have had experience with that.” In this way, one student’s personal experience became informative for his classmates’ learning.

This has happened in a Latina law professor named Camila’s classes as well. She recounts one example of a young Black male student, who has since “gone on to be a very, very successful person . . . in the corporate world,” discussing the early desegregation cases in her Constitutional Law class. While working toward the recent affirmative action cases, Camila started the class with early litigation efforts to desegregate public facilities, including swimming pools. This particular Black student “said something about, ‘You know, come on, it really all came down to white parents not wanting to have their little white girls in pools with young Black guys.’” Years later, recounting this experience as part of her DLA interview, Camila still recalls how this one student’s rationale for anti-desegregation efforts, drawing from his own experiences as a Black man, galvanized discussion in the class. She shares, “[A]ll I can tell you is that was an electrifying moment for that class, everybody in that class. Everybody in that class got it.” She makes clear how personal context helps students better understand the law as a whole: “It’s not that it elucidated a fine point of law, because there was no fine point; but all of a sudden people saw what

145. See *infra* Part III.B.2 for more on how students initiating these discussions is preferable for learning than faculty members doing so.

146. 277 A.2d 369 (N.J. 1971).

that case was about in a way that I couldn't [have explained through simply lecturing on the court opinion]." A Native American professor named Jennifer agrees, explaining, "We have a lot of very privileged white students here and when they hear their own classmates say something like, 'Well, I've been stopped, you know, [just because I am] a person of color,' [it broadens their perspective]." Regardless of how many ways she explains recent lawsuits regarding "stop and frisk," for her students "to be able to hear that in the classroom from a peer is much more powerful than if I were to make that observation."

Similarly, a Latina law professor named Marisol says, "I've been lucky that I've actually had some immigrants in class where the other students had no idea" how someone from an immigrant family would have different experiences from their own. It is especially informative when her students are forthright about their own trajectory to the United States, through legal means or not, and expose their classmates to the reasons for their sometimes perilous journeys and the challenges they encountered along the way. She says that the most powerful classroom conversations follow a student "saying, 'My parents were undocumented and I came in as an undocumented person.'" Though some students may initially believe "you're breaking the law," Marisol can see their perspective change as the conversation progresses, because "the next thing you know they're sitting next to someone who was in that experience and [learning] what that experience was like." She has also had a situation where a former refugee spoke up in class about how her "parents . . . came in as very educated and they had to mop floors" because no other jobs were available. Even "just being able to share the experience of what it's like being in the refugee camp for three years [makes] people really realize like, 'Wait a minute. [I've been] taking [a lot] for granted [and] it's not that easy.'"

Ryan, a Black law professor, shares similar experiences in both his Criminal Law and Criminal Procedure courses. There, gender diversity becomes instructive for student learning. He notes: "[W]hen I teach Criminal Law and when we talk about rape in class, the students—particularly the male students—have an opportunity to really listen to their female colleagues and understand what rape is." Especially in the current culture of ongoing campus challenges regarding sexual assault,¹⁴⁷ he notes that it is important for the men in his classroom to think more critically about "[w]hat constitutes good sexual conduct, what constitutes sexual harassment, what's crossing the line, things of that nature." Some of his female students have worked in rape crisis centers and are "willing to share the work they do with people who are suffering from these experiences." Those pieces of information "shed light on the subject and conversation,"

147. For instance, "55 colleges and universities nationwide [are] under scrutiny by the U.S. Department of Education for [their] handling of reported sex crimes on campus." See Jack Flynn, *Amherst College, Responding to Federal Title IX Probe, Cites Major Improvements in its Handling of Sexual Assault Complaints*, MASS LIVE (May 2, 2014, 10:02 AM), http://www.masslive.com/news/index.ssf/2014/05/amherst_college_responding_to.html.

especially for the male students in class who may have less context for understanding or even thinking critically about these issues. Hearing from their female classmates thus informs their own understanding of these laws.

In Ryan's Criminal Procedure course, his students of color "talk about being stopped [by the police]. They talk about racial profiling, they talk about even [being harassed by security] sometimes right outside the law school late at night." Because race in the United States has always been more than a Black-white paradigm,¹⁴⁸ true diversity matters most in terms of having a multiplicity of backgrounds represented. Ryan's Black students talk about being stopped by the police; but his Middle Eastern and South Asian students "share about their experience and why they think they got targeted for an additional screening [at the airport]." In response, the white students "are actually very open to it." They likely are curious, interested, and even recognize how understanding and applying these personal situations will help them become more familiar with the legal concepts they are studying. Thus, the white students "listen, they engage, [they] sometimes even ask questions, [like], 'Well what did you do? Do you think you did anything wrong? What do you think it was about you that made you a victim of racial targeting?'" This back and forth among the students, Ryan believes, is the best way for them to learn the law and perhaps even learn more about people who are different from themselves in the process. He says, "[I]t opens the doors for a great conversation between the two groups."¹⁴⁹

As mentioned earlier, the perspectives that diverse students bring to these classroom conversations go beyond race/ethnicity and gender. Immigrant status already has been discussed as relevant in many classes. Socioeconomic class is important, too. When asked to reflect on diversity discussions in class, a Latina professor named Carla says, "I can think of a recent one this semester. We were covering nuisance and at the time the port of Los Angeles was being sued by the local community for nuisance." She was using that ongoing lawsuit to illustrate an older one, "the *Boomer* case,¹⁵⁰ which involves cement dust, [which is now] known to cause asthma, heart [problems, and other negative] effects." One particular student contribution stood out to her and the other students in class, and stands out to her even now:

One student was from an area like that and he raised his hand and he had worked with the California legislature on this and that related to emissions and he gave an impassioned speech about the perpetuation of poverty and how this kind of pollution perpetuates poverty and it makes people sick.

148. See generally FRANK WU, *YELLOW: RACE IN AMERICA BEYOND BLACK AND WHITE* (2003); Perea, *supra* note 140.

149. When asked directly how he thinks his classes would be different if they were less diverse, Ryan says he would attempt to bring up many of those issues on his own, but does not think it would be nearly as effective because "when it comes from me, it sounds as if I'm trying to provide my own opinion or own view about police as opposed to coming from peers" whom students are more likely to believe.

150. *Boomer v. Atlantic Cement Co.*, 26 N.Y.2d 219 (1970).

He himself was physically ill from these kinds of emissions, something he had to carry, and this was an externality and we were talking about externalities.

She recalls how much this student sharing his personal and professional experience with exposure to a toxic substance helped his classmates understand the issue better, noting that “it was so right on, you know, and so honest” that his peers were bound to remember the legal rule she sought to convey through the case, recognize the importance of the case for people personally affected, and see how the law could help make injured people whole.

Similarly, a white male law professor named Adam recalls a student who had “lived outside the U.S. [who] said, ‘You know, not everyone views the United States as a world hero in world activities. I knew a lot of people [when I lived abroad] who felt like the U.S. was destroying our culture.’” Adam thought that was “an amazing statement to say.” Primarily, his class consists of “people who have lived their whole life in the U.S. and only seen one side [of this global power], which is the side that makes them safe.” Because of their American-focused perspective, Adam thought it was “really powerful” for them to also hear from someone who has seen it from another side, “a side that threatens.”

Madison, a white female law professor, has “been pleasantly surprised at how well it has worked” when she encourages students to draw on their personal experience to help illustrate legal concepts. She recalls a case from “this week in Civil Procedure” regarding a police chase involving officers who “ended up doing a maneuver to run [the suspect] off the road and he crashed into a telephone pole.” In that case, *Scott v. Harris*,¹⁵¹ the main “[legal] question is, looking at the video from the car chase, can a reasonable jury decide this officer was doing anything other than protecting public safety?” The case also touches on “other questions about who is on a reasonable jury? What kinds of life experience do they bring? Is that a decision we want judges to make?” In class, Madison “talked about the law first and then we watched the video itself. And then we had a discussion where we first looked at [it] from a legal perspective, if you’re arguing for one way on summary judgment or if you’re arguing the other way.” Thus, the focus in class was on the black letter law, understanding summary judgment, and when a judge should assume no reasonable jury could find for the nonmoving party.¹⁵² Throughout the discussion, which “was entirely volunteer, I had a tremendous amount of participation . . . from a wide range of students of different races and ethnicities as well as life experience.” Perhaps most instructive was participation from a student who was a “former police officer who now trains people in that maneuver itself, and was able to explain [how it should work in practice].” What made the conversation the most satisfying for Madison, and likely for her students, is

151. 550 U.S. 372 (2007).

152. For more on the standard for summary judgment, see FRIEDENTHAL ET AL., *CIVIL PROCEDURE: CASES & MATERIALS* (11th ed. 2013).

that “it was not only a discussion that helped [illuminate] the legal points about what summary judgment is, but I think [the students] did a really good job about articulating their own ideas about the case and policy concerns as well.” Drawing from so many personal experiences, the students are likely to better understand the legal standard and remember the law in action as well.

3. Professional and Career Advantages

“[S]tudents will become better lawyers and better serve their clients by having a more diverse law school experience.” — Carolina

A Latina law professor named Laura makes clear that there is “no downside to diversity at all.” She sees it as “enriching” the classroom environment, especially for those students who have not experienced much diversity until law school. While lauding the educational benefits of diversity, Laura sees professional benefits as the primary advantage, noting that students “won’t be able to work in the world without these experiences in law school.” She provides an example: “If you don’t learn how to interact with a woman authority figure and you leave [law school] and you find a woman judge and you never had an experience with that interaction,” the new lawyer is likely to be “marginalized” in the courtroom and beyond. She herself has “been on that end, hiring [for a prestigious position] in D.C.,” so she knows “it’s really important for people to have those experiences, because working with new lawyers, you can tell what kind of experiences they’ve had.”

A white female law professor named Jordan sees diverse classroom interactions as especially important “in a law school setting where so much is driven based on discussion and expanding your understanding or your ability to interact with others.” Especially for students who did not have many opportunities for cross-racial interaction prior to law school, “it just brings a whole phenomenal widening of your mind [that creates a] great advantage.” Ellen, a white law professor, agrees that diversity in the classroom is important because it will likely mirror students’ future workplace environments; thus, diversity is important not only because it “enhances the discussion” in the classroom, but also because “the dynamic is much more representative of the reality they will be working in,” likely in the very near future.

Preparation for the increasingly globalized profession of legal practice is a benefit that many DLA participants discuss. Michelle, a Black professor, adds that student diversity in law school is “crucial,” in large part because “it’s a global world” where it is increasingly important not only to “get along with, but [also to] actively collaborate with people who are different.” Without those skills, new lawyers “are going to be disadvantaged in the future world, that’s just the way it is.” She believes that “every elite institution these days understands that you can’t do excellence without diversity.”

A Latina named Carolina echoes many of the comments presented earlier on educational benefits, noting that student diversity “brings richness to the law school [including] a broader range of experiences, greater cultural competency, [and] training opportunities for the entire student body.” She also agrees with Michelle that as the workplace becomes “increasingly globalized,” it becomes “increasingly important to experience and value diversity.” Having diversity in many forms, “in terms of race, gender, sexual orientation, socioeconomic status, whether people come from rural or urban environments, I think all of that helps people understand that we have a broad range of experiences and we can’t assume that our experience represents everyone’s experience.” The primary benefit of engaging with diverse perspectives is that “students will become better lawyers and better serve their clients by having a more diverse law school experience.”

C. Missed Opportunities

Of course, not all participants in the DLA study hail from richly diverse law schools. Many recognize their own campuses as either relatively homogenous, or at least lacking a critical mass of students of color. In this section, law faculty members from less diverse schools lament the lack of diversity on campus, recognizing some obvious gaps in their classroom conversations, less fulfilling discussions overall, and unique challenges for the few students of color at non-diverse schools who are either silenced or risk being seen as spokespeople for their group.

1. Holes in the Conversation

“[My Black students] feel like chocolate chips in vanilla ice cream in class.” — Michael

Without diversity, many law faculty members identify a discernable gap in classroom discussion, a gaping hole in the conversation that could be filled with a critical mass of students of color.¹⁵³ John, a white male law professor, says that at his law school, “we have a minimally diverse student body, similar to our faculty.”¹⁵⁴ Instead of having a critical mass of students of color, who would likely be more comfortable adding their perspectives in class, at John’s law school, he says, “We have a handful of minority students. I wouldn’t characterize it as particularly diverse.” He sees this as a clear drawback for his students and their learning potential, noting that “the classroom experience would benefit from multiplicity of backgrounds and viewpoints and diversity.” None of the conversations that

153. The *Grutter* Court defined “critical mass” as “meaningful numbers or meaningful representation . . . that encourages underrepresented minority students to participate in the classroom and not feel isolated;” in other words, “numbers such that underrepresented minority students do not feel isolated or like spokespersons for their race.” *Grutter v. Bollinger*, 539 U.S. 306, 318–19 (2003) (internal quotation marks omitted). This Article utilizes this same definition.

154. For more on law faculty diversity and especially cross-racial interactions between law faculty members, see Deo, *supra* note 96, at 53–56.

characterize the richly diverse classes discussed earlier in this Article (for example, covering stop and frisk or the reasonable person) occur with regard to race in John's classes.

Similarly, a white male law professor named Ian notices the drawbacks of a lack of diversity in his class of primarily white students when discussing "the affirmative action cases." The holes in the conversation are often obvious, as when one of his white students "very self-consciously celebrates the importance of [educational diversity, saying,] 'You know, I'm from rural Montana. I did not encounter a Black student or Black person until I went to college,' [talking] very from the heart." While Ian appreciates the student's candor, he also wonders, "[W]hat do you do with that?" He is in a teaching situation "where there may not even be [any] or only one or two Black students at all in the room." While a critical mass of Black students could likely contribute to a healthy and vigorous debate about affirmative action, whether there are zero, one, or two African Americans in class, there is likely to be little conversation even if a classmate brings it up in a positive way. This is because when there are only one or two members of an identity group present, they may be uncomfortable speaking up.¹⁵⁵ Their silencing is thus akin to there being no minorities present at all.¹⁵⁶ As a result, the students who have had little exposure to meaningful cross-racial interaction before law school will remain disadvantaged because their few classmates of color are often uncomfortable speaking up in a setting where they are so singled out.

Michael, a Black law professor, speaks to this directly. He knows firsthand what it is like to be the only Black man in a room full of others. Thus, he identifies with his students of color who are uncomfortable speaking up in class, even when their perspectives would be valued or appreciated by their classmates. Michael acknowledges that it would help to have "a critical mass [for them to] feel more comfortable. They wouldn't feel like chocolate chips in vanilla ice cream in class." Standing out in such an obvious way, starkly visible in contrast to the traditional white students all around them, many students of color thus do not voice their perspectives; the unfortunate result is that everyone misses the opportunity to learn from one another.¹⁵⁷

155. For more on how tokenization can lead to silencing, see *infra* Part III.C.3.

156. The *Grutter* Court specifically noted the challenges associated with speaking up in class for members of underrepresented groups; it determined that a critical mass of students would add sufficient "numbers such that underrepresented minority students do not feel isolated or like spokespersons for their race." *Grutter*, 539 U.S. at 319 (internal quotation marks omitted).

157. For more on tokenization and silencing, see *infra* Part III.C.3, especially Natalie's example of her one Black student in class.

2. Less Fulfilling Discussions

“[I]t was just like you were teaching to somebody on Mars about race issues.” — Martha

In addition to there being fewer perspectives shared in class, a non-diverse classroom also creates less fulfilling conversations overall. Natalie, a multiracial professor, sums it up succinctly: “[Y]ou cannot have a rich discussion of the way Criminal Law works if there are no Black people [present].” Camila, a Latina, notes this problem in her classes as well, stating, “It has been difficult for me to think of having a meaningful conversation about race or gender in a non-diverse classroom.” Similarly, a Native American law professor named Erin recognizes that classes composed primarily—or perhaps exclusively—of privileged whites prevents meaningful conversations about numerous important legal topics. She notes:

I don’t think I had anybody that was racially diverse in the class. And that kind of made it really difficult when we got to the discussion about environmental justice because not only was it that they weren’t necessarily diverse, I got this sense that the majority of the students in the class really came from privileged backgrounds. And, so it was really difficult for them to understand, kind of, the impacts of pollution on communities when the communities don’t really have other options. So I think our discussion of environmental justice was not as full or interesting as a result. I certainly didn’t enjoy it as much.

The lack of diversity of racial/ethnic background as well as experience made for a relatively bland classroom discussion in Erin’s class, which was “not as full or interesting” as it could have been with more perspectives included.

In one particular school, a Latina professor named Martha “felt it was very tough . . . to teach Constitutional Law” because her students were “a bunch of white kids” with very limited exposure to people of color. When pressed on the challenges associated with teaching this particularly contextually dependent area of law to a racially homogenous group, Martha explains that her students “just had no appreciation of race. I mean it was just like you were teaching to somebody on Mars about race issues [A]bsolutely it makes a big difference to have a racially diverse classroom.”

Isabella, a white female professor, is more specific about what has been missing from her classes. She chooses religion in her First Amendment class as an example of a course that blooms when filled with students from diverse religious backgrounds, but falls flat when all students share a similar religious background. Isabella says that especially in “First Amendment, it would be nice to have more religious diversity, maybe more Jewish, Muslim, and certainly the Eastern religions that I don’t see represented frequently.” The reason that religious representation matters, and why it is especially important to have minority religions represented, is to show students that their perspective is not the only one out there—a key

concept for budding lawyers to master. Isabella states: "I always make a joke that 97 percent doesn't see any harm in issues that deal with religious minorities because you're in the 97 percent. In my classroom, if I've got everybody in the 97 percent how can I help them understand?" Again, having just one or two religious minorities does not tend to improve the debate because those token individuals feel the pressure of being the "other." Isabella continues:

And if there's only one person in the minority religion that's problematic because that's not representative and might become a stereotype. If you had more Muslims, you would have more diversity. If you had more African American individuals, you'd have more diversity, but having small samples leads to stereotypical thinking and groupthink.

3. Being a Spokesperson

"If you don't have that critical mass, I think it makes the few diverse students you have feel uncomfortable." — Jordan

In addition to the many ways in which a lack of diversity creates missed opportunities for all students in the classroom, the DLA data make clear that it also clouds the experience of the few students of color in class. This finding, from the faculty perspective, triangulates existing data in this area showing that when there are only a few students of color in any given class, they tend to be alienated and sometimes disengage from learning.¹⁵⁸ Existing literature documenting the student perspective shows that when there are only limited numbers of students of color in a class, they are especially unlikely to speak up for fear that their comments will be taken as representing not their individual perspective, but instead the perspective of the group their peers believe they belong to.¹⁵⁹

In fact, the DLA data reinforce studies from the student perspective, documenting how faculty also recognize student alienation, disengagement, and especially an unwillingness to be seen as the spokesperson for their race when they are one of the few in class from a particular background. A white female law professor named Ellen notes that her best classes are those where there is a critical mass of students of color, so that they themselves can engage in meaningful dialogue and even disagreement with one another. In this way, they have an opportunity to demonstrate their individuality rather than being perceived as representing their group. Ellen says, "I like looking out over a large classroom and knowing that if a student of color speaks up, there are enough other students of color that they're not designated as the spokesperson for their race."

158. See, e.g., Deo, *supra* note 81, at 15.

159. It is altogether too common to confront "alienation and isolation for the few students of color who attend predominantly white schools, whether in high school, college, or beyond." *Id.* at 18.

Many professors do not share Ellen's experience, even if they would like the high level of diversity she enjoys.¹⁶⁰ A white law professor named Marybeth acknowledges that a lack of diversity "can be a problem [especially] when you're talking about issues of diversity and race" and there are no or few students of color in class. When that happens, the white students in class are essentially "talking about an 'other' who's not in the class." Perhaps even more problematic is "when you have [only] one African American student, which happens many times; it's like the minute you bring up a race issue that person becomes the spokesperson." Though Marybeth is rarely in a situation where she is the racial minority in the room full of "others," she is nevertheless attuned to the alienation accompanying this situation for her students of color. She notes, "I feel very uncomfortable with that because it's not a position I would want to be in."

A white professor named Jordan agrees, emphasizing how tokenism leads to silencing in the classroom: "If you don't have that critical mass, I think it makes the few diverse students you have feel uncomfortable; maybe they are not going to voice opinions." Instead, her goal would be "to have that critical mass where everyone feels comfortable, and it enables everyone to broaden the discussion and bring a variety of perspectives to the classroom conversation." Natalie, a multiracial law professor, recalls a disappointing class discussion about criminal defendants that ended with "[the] one Black guy in the room" approaching her after class to say, "'Professor, I disagree with what everybody says but I don't want to be [seen as] The Black Guy.'" She told the student immediately that she understood and, in order to not single him out, did not call on him during challenging classroom conversations covering poignant cases involving race. She told him specifically, "'I just don't want you to be the representative of your race and the demographics of our law school have often put you in that position.'"

Ian, a white male law professor, suggests that having a meaningful number of students of color in class is connected to the likelihood of having valuable conversations that draw from personal background, "which goes to show that there is something to the critical mass discussion in *Grutter*."¹⁶¹ Ian personally enjoys a broad range of diversity in class, not only along racial/ethnic lines, noting, "I've had students every year that have spoken based on their background—whether Native American, African American, and mainly white military, veteran status—all of those perspectives are helpful." Yet, especially when issues of race are central to the case being discussed, "in teaching Con Law and talking about affirmative action, for example [and especially because we are] a public law school," meaningful conversations are less likely to occur because "there simply is not always a critical mass of certainly African American students, Latino students, and

160. See *supra* Table 5 and accompanying discussion (analyzing how most law professors would prefer greater diversity on campus than they currently encounter).

161. Ian is referring here to the Court's point in *Grutter* regarding critical mass, discussed *supra* note 153.

other such groups.” Ian sees the importance of including as many perspectives as possible and knows it is more likely that individuals will speak up and be seen as representing their own opinion—not their group’s opinion—when there is a critical mass; yet:

The idea of the critical mass—of having enough people in there so that no one feels like they are speaking for their group, whether its veteran, whether its environmentalist, or whether it’s a Native American student—is not something we can succeed at all the time given the demographics of our state and our school.

However, through affirmative action, at least he and his colleagues have the opportunity to teach a diverse group; the admissions officers at his school can take account of race when making decisions about the overall student population at the school because they can legally rely on affirmative action to help diversify the student body.

CONCLUSION

Affirmative action remains under attack. So does educational diversity, the only remaining non-remedial compelling state interest that courts have sanctioned to justify the consideration of race in higher education admissions decisions.¹⁶² While the Supreme Court has accepted educational diversity as a permissible goal for institutions of higher education to consider when determining the composition of their student body, the goal stands on shaky legal ground.¹⁶³ Empirical research should be used to bolster educational diversity in this time of need. The vast majority of research on educational diversity in higher education centers on the student experience and perspective.¹⁶⁴ Yet, faculty members may have the most useful contributions regarding the actual educational and career benefits of educational diversity, given their unique roles in the classroom as both teachers of substantive material and facilitators of relevant discussions.

This Article presents findings from the DLA project, revealing law faculty members’ insights into student diversity. Relying on law faculty experiences, it is clear that student diversity plays an important role in the legal education process. Law faculty members from all racial/ethnic backgrounds note the various ways in which student diversity is an integral part of student learning. For instance, many say that a diverse student body means that there are multiple perspectives in every classroom and an opportunity to draw from personal context to illuminate what is otherwise often abstract black letter law. Relevant context initiated and discussed by students is preferred to faculty lecturing on it, as students respond better to their own peers and colleagues raising personal experiences and observations. There are also myriad career advantages associated with educational diversity, for future lawyers who will be serving clients from

162. Deo, *supra* note 8, at 664.

163. *Id.* at 662.

164. See, e.g., Deo et al., *supra* note 70; Gurin et al., *supra* note 70; Deo, *supra* note 3.

diverse backgrounds, those who might work internationally, and for any one of them who may one day come before a female judge.

The flip side of the benefits of educational diversity involves the drawbacks associated with a lack of diversity. A number of participants in the DLA study have experience with low student diversity and lament its many negative attributes and results. When no or few diverse students are present, there are obvious gaps in classroom conversations that law faculty members are not always equipped or able to fill. Many professors have lectured through the discomfort of covering police brutality, racial profiling, and deportation-as-punishment without African Americans, Muslims, or Latinos present. When the classroom lacks a critical mass of diverse students, many DLA participants note that the conversations as a whole tend to be bland rather than robust. Furthermore, many law faculty members recognize what students themselves have noted as research subjects participating in empirical research for years: when there is nobody else from their background in the classroom, students of color tend to carry the burden of representing their race, where their classmates expect them to be a spokesperson for their race. Perhaps for all these reasons, the quantitative data reveal that law faculty members from all racial/ethnic backgrounds and men and women alike tend to strongly support diversity in the classroom.

Educators, administrators, and admissions officers surely can rely on these findings to continue supporting diversity efforts at institutions around the country. When over 75 percent of law faculty from all backgrounds—including whites and nonwhites as well as men and women—support greater diversity on campus, it is clear that even greater efforts are needed to achieve it. Ironically, the courts may be gearing up to turn away from educational diversity altogether. Armed with this detailed empirical analysis from law faculty members, those fighting to preserve affirmative action and educational diversity may now be better equipped to detail to the courts not only the many benefits of educational diversity, but also the challenges and drawbacks associated with a lack of diversity. Without educational diversity as a compelling state interest, we may lose affirmative action altogether.¹⁶⁵ And without affirmative action, we may all be at institutions where racial/ethnic homogeneity impedes teaching and learning for us all.

165. For more on the legal doctrine associated with affirmative action, see Part I.A. See also Deo, *supra* note 8, at 668–73.

Women are underrepresented in fields where success is believed to require brilliance

Meredith Meyer^{1*}, Andrei Cimpian² and Sarah-Jane Leslie³

¹ Department of Psychology, Otterbein University, Westerville, OH, USA, ² Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA, ³ Department of Philosophy, Princeton University, Princeton, NJ, USA

OPEN ACCESS

Edited by:

Stephen J. Ceci,
Cornell University, USA

Reviewed by:

Sapna Cheryan,
University of Washington, USA
Elizabeth A. Gunderson,
Temple University, USA

*Correspondence:

Meredith Meyer,
Otterbein University,
1 South Grove Street,
Westerville, OH 43081, USA
mmeyer@otterbein.edu

Specialty section:

This article was submitted to
Developmental Psychology, a section
of the journal *Frontiers in Psychology*

Received: 01 January 2015

Paper pending published:
25 January 2015

Accepted: 15 February 2015

Published: 11 March 2015

Citation:

Meyer M, Cimpian A and Leslie S-J
(2015) Women are underrepresented
in fields where success is believed to
require brilliance.
Front. Psychol. 6:235.
doi: 10.3389/fpsyg.2015.00235

Women's underrepresentation in science, technology, engineering, and mathematics (STEM) fields is a prominent concern in our society and many others. Closer inspection of this phenomenon reveals a more nuanced picture, however, with women achieving parity with men at the Ph.D. level in certain STEM fields, while also being underrepresented in some *non*-STEM fields. It is important to consider and provide an account of this field-by-field variability. The field-specific ability beliefs (FAB) hypothesis aims to provide such an account, proposing that women are likely to be underrepresented in fields thought to require raw intellectual talent—a sort of talent that women are stereotyped to possess less of than men. In two studies, we provide evidence for the FAB hypothesis, demonstrating that the academic fields believed by laypeople to require brilliance are also the fields with lower female representation. We also found that the FABs of participants with college-level exposure to a field were more predictive of its female representation than those of participants without college exposure, presumably because the former beliefs mirror more closely those of the field's practitioners (the direct “gatekeepers”). Moreover, the FABs of participants with college exposure to a field predicted the magnitude of the field's gender gap above and beyond their beliefs about the level of mathematical and verbal skills required. Finally, we found that beliefs about the importance of brilliance to success in a field may predict its female representation in part by fostering the impression that the field demands solitary work and competition with others. These results suggest new solutions for enhancing diversity within STEM and across the academic spectrum.

Keywords: gender, stem, lay theories of success, field-specific ability beliefs, diversity in academia

Introduction

A recent article in *Scientific American Mind* begins: “Try this simple thought experiment. Name 10 female geniuses from any period of history. Odds are you ran out of names pretty quickly” (Upson and Friedman, 2012, p. 63). The thought experiment can be adapted: try to name 10 female figures in popular culture who—like Sherlock Holmes, Dr. House, or Will Hunting—are characterized by their innate brilliance, their raw intellectual firepower. As before, one rapidly runs out of names. Whatever the cause, the message is clear: women are not culturally associated with such inherent gifts of genius (Bennett, 1996, 1997, 2000; Tiedemann, 2000; Rammstedt and Rammsayer, 2002; Furnham et al., 2006; Kirkcaldy et al., 2007; Upson and Friedman, 2012; Lecklider, 2013;

Stephens-Davidowitz, 2014). The consequences of this stereotype are likely wide-ranging. In the current study, we focus on one of these consequences, asking whether such a pervasive cultural message might have a role in shaping individuals' academic and career paths. Specifically, if it is widely believed that men tend to possess more intellectual ability than women, then women may be discouraged from entering into fields that are thought to require this ability. We call this the field-specific ability beliefs (FAB) hypothesis (**Figure 1**): *the more a field is believed to require raw brilliance, the fewer the women* (Leslie and Cimpian et al., 2015). We test this hypothesis in the context of gender gaps in academia, investigating whether these gaps are predicted by how much laypeople assume that success in various fields rests on raw ability.

Gender disparity in academia has been a generative topic of research for many years, with contemporary focus on this issue largely centering on men's and women's participation in (natural) sciences, technology, engineering, and mathematics (STEM). The general phenomenon is clear: on average, female representation in STEM fields (particularly those that are math-intensive) is lower than in the social sciences and humanities (SocSci/Hum). Though the magnitude of this gap has largely decreased across the last several decades, the difference is still reliable, prompting a number of efforts to explain it (for reviews, see Ceci and Williams, 2007; Ceci et al., 2009, 2014; Hill et al., 2010).

The low number of women in STEM is indeed of real concern. However, it is also important to observe that there is at least as much variation in female representation *within* STEM and SocSci/Hum as there is *between* them. For instance, when examining the number of recent doctoral degrees earned by women in the U.S. in 30 different fields (**Table 1**), STEM fields are characterized by female representation ranging from

just under 20% (physics) to over 50% (molecular biology; National Science Foundation [NSF], 2011). An even larger range is observed within SocSci/Hum fields, with women earning fewer than 35% of doctoral degrees in philosophy and economics, yet over 75% in art history. Indeed, the range of variation is so wide that many STEM fields feature *higher* female representation at the Ph.D. level than many SocSci/Hum fields. Given this large variation within STEM and SocSci/Hum considered separately, it is apparent that expanding the focus of inquiry to include gender gaps in both STEM and SocSci/Hum might provide new insights into the problem of female underrepresentation. In the current study, we adopt such a broad focus, examining whether the FAB hypothesis can account for the field-by-field variability observed across the entire academic spectrum.

Initial Evidence for the FAB Hypothesis

In a recent study, we sought to test whether the FABs held by *academics* could predict the wide field-by-field variability observed in female representation across both STEM and non-STEM fields (Leslie and Cimpian et al., 2015). We gathered responses from a sample of over 1800 professors, graduate students, and post-doctoral researchers from research-intensive universities across the U.S. in 30 different fields (12 STEM, 18 SocSci/Hum; **Table 1**). We first asked participants to report on their beliefs regarding what was required for success in their own field, focusing on assessing beliefs about the relative importance of intrinsic, stable ability vs. effort and practice (see Dweck, 2000, 2006). We then used these items to provide a metric of field-level ability beliefs; each field received a FAB score expressing average endorsement of ability vs. effort across individuals within a given field (with higher scores indicating more emphasis on raw ability). Results included three important findings. First, FABs

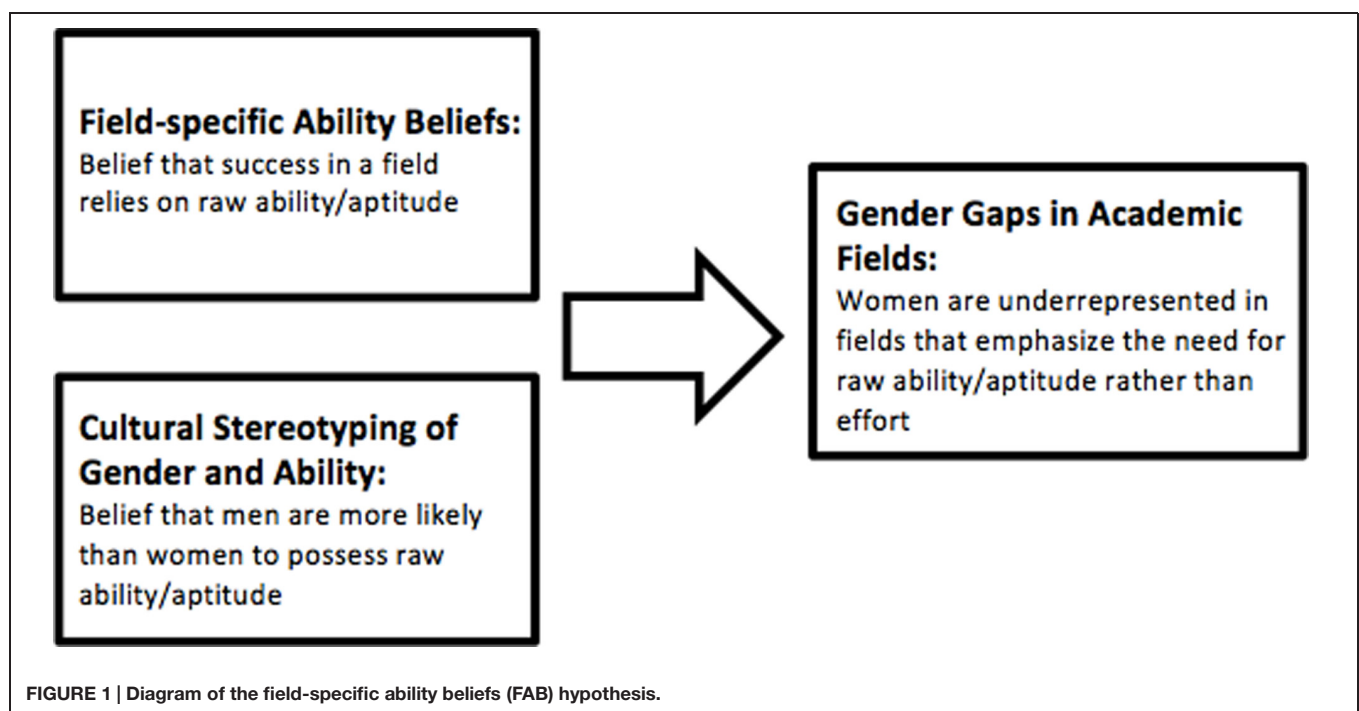


TABLE 1 | Percent of American Ph.D.'s earned by women in 2011* in science, technology, engineering, and mathematics (STEM) and Social science/Humanities fields.

STEM Field	% of Ph.D.'s who are Female	Social science/Humanities field	% of Ph.D.'s who are Female
Physics	18.0	Music theory and composition	15.8
Computer science	18.6	Philosophy	31.4
Engineering	22.2	Economics	34.4
Mathematics	28.6	Middle Eastern studies	38.1
Astronomy	29.2	Classics	41.8
Earth sciences	36.2	Political science	43.1
Chemistry	37.8	History	45.0
Statistics	41.6	Archeology	52.3
Biochemistry	45.4	Linguistics	59.2
Neuroscience	49.4	Anthropology	59.6
Evolutionary biology	49.8	Spanish/Spanish literature	59.9
Molecular biology	54.4	Comparative literature	60.9
		Sociology	61.3
		English literature	62.4
		Communication studies	64.2
		Education	69.3
		Psychology	72.1
		Art history	76.8

*Data from 2011 NSF Survey of Earned Doctorates.

were strongly negatively associated with female representation (as measured by proportion of U.S. Ph.D. degrees earned by women; National Science Foundation [NSF], 2011), providing initial broad support for the hypothesis: there were fewer women in fields believed to require stable, raw talent. Second, ability beliefs were predictive of female representation over and above whether a field was STEM or SocSci/Hum, suggesting that the FAB hypothesis can account well for the wide variability observed even within the two categories of fields. Finally, the FAB hypothesis outperformed a number of other constructs often theorized to contribute to gender gaps in academia, including field-specific variation in work-life balance (e.g., Ferriman et al., 2009), selectivity (e.g., Hedges and Nowell, 1995), and reliance on skills related to systemizing vs. empathizing (e.g., Baron-Cohen, 2002).

The Present Research

Do Laypeople's Beliefs Predict Female Representation?

Results from Leslie and Cimpian et al. (2015) suggest that women are underrepresented in fields whose practitioners consistently endorse the idea that success rests on brilliance. The current study extends existing work on the FAB hypothesis by exploring what *non-academics* believe is required for success in a variety of fields. We hypothesize that the FABs endorsed by academics will be shared, at least to some extent, by people outside of academia as well. This is an important extension of the FAB framework because, to the extent that these beliefs are shared by the general public, they could influence women's career choices in a much broader variety of contexts than the beliefs of academics *per se*. If these beliefs pervade our society, then—in combination with the stereotypes against women's intellectual abilities—they could lead a variety of individuals (parents, teachers, peers, etc.) to see women as somewhat unsuited for “brilliance-required”

domains. Even in the absence of such biased treatment, widely shared ability beliefs similar to those previously identified in academics could lead young women to doubt that they could succeed in brilliance-focused disciplines and thus to decide against pursuing careers in them. Our main prediction is thus that laypeople's beliefs, like those of academics, will predict female representation: the more a field is believed to require intellectual brilliance, the fewer the women.

We can also formulate a more detailed hypothesis here: people with more exposure to the fields in question (e.g., via college classes) will have FABs that predict female representation more precisely than the beliefs of people with less exposure. We expect this to be the case because the FABs of those with more exposure to a discipline will likely be more similar to those of the practitioners of that discipline, and are thus more likely to be similar to the kinds of beliefs that students will encounter and absorb as they start to consider higher education and careers in these fields (Leslie and Cimpian et al., 2015). In Study 1, we tested the prediction that college-exposed individuals' ability beliefs would better predict gender gaps in representation by first dividing our participants into those who had college exposure to a field and those who did not, and then exploring whether the beliefs of the college-exposed group predict female representation at a more fine-grained level.

A final aim of Study 1 was to address an alternative explanation for the hypothesized relationship between ability beliefs and female representation. As we have argued, underlying the main predictions described above is our claim that FABs influence women's academic and career choices. However, might laypeople's beliefs be simply inferred from their pre-existing knowledge about the proportion of women in the different fields? For instance, our participants—particularly those who have had

college experience in the relevant fields and thus have had the opportunity to witness gender disparities firsthand—might rely on stereotypes against women’s intellectual abilities to arrive at the conclusion that fields with few women must require high levels of such abilities, whereas fields with many women must not. To address this possibility, we asked participants to estimate the proportion of women in the fields under investigation, and assessed whether participants’ ability beliefs still predicted female representation independent from these estimates. If so, this would undermine the possibility that participants simply infer their ability beliefs from their estimates of the field’s diversity.

Assessing Beliefs Beyond FAB

There are surely many other dimensions that vary among fields and influence the gender breakdown of the people who participate in them, and we do not claim that FAB is the only factor in determining academic gender gaps. Indeed, as we observed at the outset, other such factors have been evaluated extensively in prior studies (for reviews, see Ceci and Williams, 2007; Ceci et al., 2009, 2014; Hill et al., 2010). Although an exhaustive evaluation of these additional factors is outside the scope of the current studies, we take up this issue within the framework of evaluating other *beliefs* about what is required for success in the fields under investigation. In particular, Study 2 examines two questions. First, do field-specific beliefs about the importance of intellectual brilliance reduce to beliefs about specific types of skills required for success? Specifically, do they reduce to beliefs about the degree to which mathematical and verbal skills are required for individual fields? Second, is the relationship between FABs and female representation mediated by beliefs about what kinds of work (solitary vs. collaborative; competitive vs. cooperative) are required?

Our first question addresses a potential alternative explanation for the predictive power of FABs. A critic might note that the extent to which mathematics is involved in a field appears to be particularly predictive of whether women are underrepresented or not: fields that are math-intensive attract and retain fewer women, with math-intensive STEM fields (e.g., engineering, math, or physics) characterized by the most extreme gender disparities (in comparison to STEM fields that are less math-intensive, like the life sciences, which often feature parity or even a predominance of women; National Science Foundation [NSF], 2011). The smaller number of women in math-intensive fields may be due in part to the cultural belief that math is “for” males, a belief that appears to emerge as early as elementary school and may contribute to women’s reduced interest in careers that require it (Fredericks and Eccles, 2002; Herbert and Stipek, 2005; Cvencek et al., 2011). In light of this evidence, one might ask: is it possible that the “intellectual brilliance” at the heart of the FAB hypothesis is just another way of referring to mathematical aptitude, which is also popularly conceived as a fixed, innate quantity? That is, might it be the case that people’s FABs simply reduce to their beliefs about how much individual fields require math over other kinds of skills (e.g., verbal skills)?

Results from Study 1 could bear on this question. If, as hypothesized we find that FABs are capable of predicting female representation across a variety of fields, including those unlikely

to be thought of as drawing on mathematical skills (such as most social sciences and humanities disciplines), it is unlikely that these beliefs are *merely* capturing people’s beliefs about field-specific mathematical requirements. However, it is important to more directly establish whether FABs are distinct from beliefs about the importance of mathematics. To do so, in Study 2 we tested whether beliefs about raw ability and brilliance predict unique variance in gender gaps, beyond that predicted by people’s beliefs about how much individual fields rely on mathematical and verbal ability.

Finally, it is worth noting that, as in Study 1, college exposure may matter. Participants with college experience likely have more nuanced, differentiated beliefs both about which fields require mathematical skills and about which fields require intrinsic ability. We thus hypothesized that when looking specifically at individuals with college exposure, FABs would independently predict female representation over and above beliefs about math and verbal skills, supporting the idea that FAB can account for female representation across the academic spectrum.

Next, we turn to the issue of potential beliefs that may mediate the relationship between FABs and female representation. In particular, we explore the possibility that people’s beliefs about the importance of brilliance vs. effort for success in a field give rise to differentiated perceptions of the kind of atmosphere that field promotes. We focused our exploration on two important aspects of a field’s atmosphere that (1) could be plausibly inferred based on the field’s presumed emphasis on brilliance, and that (2) men and women have diverging attitudes toward: namely, the extent to which the field requires *competition* (vs. collaboration) and *solitary work* (vs. group work; e.g., Lippa, 1998; Niederle and Vesterlund, 2007; Diekmann et al., 2010; Gupta et al., 2013; Lippa et al., 2014). There are several reasons why “brilliance-required” fields might also be presumed to require competition and solitary work. If a field values intellectual prowess, it is reasonable to expect that it would also encourage *displays* of that sort of ability, which might in turn encourage competition between individual practitioners. After all, it is only by comparing one’s ability against others (by participating in contests, engaging in aggressive debates, being harshly critical of others’ perceived mistakes, etc.) that one can reveal how brightly one’s intellectual ability shines. Working with others in cooperative contexts, on the other hand, would make it hard to assess whose talent was responsible for any ultimate success attained, so this type of collaborative work may be assumed to be rare within fields that prize brilliance. The inference that brilliance-requiring fields involve solitary, and often competitive, work is also likely to be supported by pervasive cultural tropes that portray brilliance and genius as qualities that a person possesses and displays in isolation rather than as part of a team of collaborators (e.g., Shenk, 2014). In turn, these inferences about the nature of the work environment in a field may influence whether young men and women consider careers in it because males and females are socialized to place different value on communal vs. agentic goals and on collaborative vs. competitive interactions. In other words, the downstream inferences licensed by FABs may be part of the reason why these

beliefs are predictive of gender gaps¹. We tested this hypothesis in Study 2.

Summary of Predictions

Study 1 examined two main predictions, one broad and one more specific. Broadly speaking, we expected that there would be a relationship between laypeople's FABs and female representation, such that fields believed to require brilliance would have fewer women. At a greater level of specificity, we expected that college exposure would differentiate the predictive power of FABs, such that the beliefs of those exposed to the fields during college would be particularly predictive. Finally, Study 1 also examined whether ability beliefs independently predicted female representation above and beyond people's *estimates* of female representation (suggesting that any observed relationship between ability beliefs and actual female representation did not emerge simply because individuals constructed FABs from their beliefs about female representation).

Study 2 was designed to replicate the main findings of Study 1, and to extend the inquiry into additional beliefs that might relate to gender disparities. We made two predictions. First, we predicted that FABs would *not* reduce to people's beliefs about mathematical skill, particularly when examining beliefs from individuals with college exposure in the field. Second, fields that are believed to require raw ability should also be perceived as requiring solo work and competition; in turn, these perceptions should predict gender gaps, with fewer women obtaining Ph.D.'s in fields assumed to demand high levels of solo work and competition. In other words, we expected that beliefs about solo work and competition would mediate (at least partially) the observed association between FAB and gender breakdowns.

Study 1

Method

Participants

Participants included 307 individuals recruited via Amazon's Mechanical Turk (MTurk), an online crowd-sourcing platform^{2,3}. Only participants reporting themselves as living in the U.S. and with prior MTurk approval rates of 90%

or above were included. Participants were compensated \$0.75 for survey completion. Data were excluded from an additional 48 individuals who (1) failed to complete the survey, (2) answered an attention-check question incorrectly, (3) had IP addresses indicating they were outside the U.S., and/or (4) had IP addresses indicating they had completed similar studies in the past.

Materials and Procedure

To avoid participant fatigue, we created three versions of the survey, each of which contained 10 of the 30 fields under investigation. (Fields were identical to those examined in our original study of academics (Leslie and Cimpian et al., 2015)). Fields were chosen to represent a broad spectrum of social sciences, humanities, and STEM disciplines. Approximately equal numbers of subjects participated in the three versions, and assignment was random (Version 1, $n = 103$; Version 2, $n = 101$; Version 3, $n = 103$). Each version included three humanities subjects, three social science subjects, and four STEM subjects. Each survey contained four questions assessing FABs about each of the 10 fields (from Leslie and Cimpian et al., 2015; **Table 2**). Questions were presented individually in random order with all 10 fields listed beneath each question. Participants indicated their agreement with the statement as it applied to each field using a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree, with eight as an option to indicate "don't know"). Two attention-check questions were also included to ensure that participants were attending to the task.

Next, a series of questions asked about participants' academic exposure to the 10 fields, including whether they had had (1) a high school class, (2) a college class, and/or (3) a graduate-level class in each of them. Participants were also asked to estimate how many women had received American doctoral degrees in each field in the recent past, with 10 response options corresponding to 10% intervals ranging from 0 to 100%. A final set of questions asked about demographic information (gender, age, ethnicity, and race).

For each field, we calculated FAB scores by averaging scores across participants from the four ability belief questions. Higher scores indicated more emphasis on brilliance. Three separate FAB scores were calculated: (1) All Participants' FAB (using data from all participants except those with graduate level experience in the field)⁴, (2) College Exposure FAB (using data from participants who had taken college, but not graduate level, courses in the field) and (3) No College Exposure FAB (using data from participants who had taken neither college nor graduate courses in the field). The four items had high internal reliability (for all participants, $\alpha = 0.90$; for College Exposure, $\alpha = 0.93$; for No College Exposure, $\alpha = 0.89$).

¹We acknowledge that beliefs about solo/competitive work mediating FAB's relationship with women's representation represents only one possible causal pathway; it is also possible that people could perceive the solo/competitive nature of a field and then conclude that it requires raw ability. More generally, we also note that there are likely many more factors involved in the pathways that ultimately result in the observed field-by-field variation in women's representation. More comprehensive exploration of these factors, as well as experimental work, will be needed to definitively establish how FABs influence the observed gender gaps.

²All human subjects research reported in this paper was approved by the Institutional Review Board of the first author's home institution.

³Mechanical Turk offers a convenience sample rather than a fully nationally representative sample. Analyses of American MTurk workers have demonstrated that women are overrepresented, that workers are typically younger and more educated than average, and that Blacks and Hispanics are underrepresented (Berinsky et al., 2012; Paolacci and Chandler, 2014). Thus, we do not claim to be capturing beliefs that are fully representative of the U.S. public. Nevertheless, the diversity of an MTurk sample is arguably higher than that of most samples used in human subjects research (i.e., college samples), and it provides a good source of data for an examination of beliefs held by individuals outside of academia.

⁴We excluded data regarding individual fields if they were provided by people reporting graduate-level experience in that field. We did so because we wanted to exclude beliefs held by people with extensive familiarity with the field gained through graduate-level exposure, allowing the focus of the current study to be restricted only to individuals with no college experience vs. college experience with the field.

TABLE 2 | Survey items for Study 1 and Study 2.**Field-specific ability beliefs**

- Being a top scholar of [field] requires a special aptitude that just can't be taught.
- If you want to succeed in [field], hard work alone just won't cut it; you need to have an innate gift or talent.
- With the right amount of effort and dedication, anyone can become a top scholar in [field]. (R)
- When it comes to [field], the most important factors for success are motivation and sustained effort; raw ability is secondary. (R)
- To succeed in [field] you have to be a special kind of person; not just anyone can be successful in it. (in Study 2 only.)
- People who are successful in [field] are very different from ordinary people. (in Study 2 only.)

Estimate of female representation (Study 1)

Please provide your best guess or estimate to this question: in the recent past, what percentage of doctoral (Ph.D.) degrees from American universities do you think have been earned by women in [field]?

Verbal and mathematical ability (Study 2)

- Top-level success in [field] depends to a large extent on one's verbal ability.
- Top-level success in [field] depends to a large extent on one's mathematical ability.

Solo and competitive work (Study 2)

- [Field] is a field in which you spend a lot of time working by yourself rather than being around other people.
- [Field] is a field in which competition with others is much more common than collaboration.

(R) indicates items that were reverse scored.

Responses to all items except estimate of female representation were given on a 7-point scale (1 = strongly disagree to 7 = strongly agree), with an additional option for "don't know." Responses for estimate of female representation were given on a 10-point scale, with each point representing a 10% increment.

Results and Discussion

Study 1 tested two main predictions. First, we expected that participants' FABs would be correlated with female representation regardless of participants' level of direct prior exposure with the fields (via courses). Second, we predicted that beliefs held by individuals with college experience would nevertheless be predictive of female representation at a *finer-grained level* than those of people with no college experience. In particular, we expected that the College Exposure, but not the No College Exposure, FAB scores would predict female representation even after taking into account a gross STEM vs. non-STEM distinction between fields, which would speak to the ability of the College Exposure FAB scores to predict the complex field-by-field variability in female representation observed within these broad domains. Finally, we examined whether beliefs of college-exposed and non-college-exposed individuals predicted actual female representation independent of participants' *estimates* of female representation. If so, this would rule out the possibility that ability beliefs predicted female representation for the trivial reason that they were inferred from participants' pre-existing knowledge about gender disparities.

To assess our first prediction, we examined the correlation between FABs and female representation. Any fields for which we received fewer than 10 participants in either the no-college-experience or college-experience samples were removed from the analysis; estimates based on so few participants would likely be unreliable. This resulted in 29 fields being retained for analysis. (The single removed field was neuroscience; only seven individuals reported college experience with this field.) As predicted, fields believed to require brilliance had lower female representation, $r(27) = -0.59, p = 0.001$ (**Figure 2**).

To address the second prediction, we separately examined beliefs held by people with college exposure and those held by people without college exposure. Beliefs of both groups were significantly negatively associated with female representation:

College Exposure scores, $r(27) = -0.67, p < 0.001$, and No College Exposure scores, $r(27) = -0.51, p = 0.005$. Steiger's z score comparison (Lee and Preacher, 2013) indicated that College Exposure scores were more strongly associated with female representation than No College Exposure scores, $z = 2.09, p = 0.037$, providing initial support for the prediction that college-exposed individuals' beliefs would relate more strongly with representation. We then investigated whether the ability beliefs of these two groups predicted female representation above and beyond whether a field was STEM vs. a social science/humanities discipline (i.e., non-STEM). Two separate multiple regression analyses were performed with female representation as the dependent variable and two predictors: a STEM/non-STEM indicator variable and either (1) College Exposure FAB scores or (2) No College Exposure FAB scores. These analyses indicated that, as hypothesized, the FABs held by participants with college exposure to the fields were uniquely predictive of female representation, above and beyond whether the fields were in STEM or SocSci/Hum ($\beta = -0.44$, bootstrapped $p = 0.013$), whereas the beliefs of participants without college exposure were not ($\beta = -0.15$, bootstrapped $p = 0.449$).

Finally, we added college-exposed and non-college-exposed participants' estimates of female representation as predictors to the two regressions above. Consistent with our argument, the FABs of college-exposed participants remained a significant predictor of actual female representation even when adjusting for these participants' estimates of female representation ($\beta = -0.41$, bootstrapped $p = 0.043$; see **Table 3**). In contrast, the beliefs of participants without college exposure were not a significant predictor of female representation in this model ($\beta = -0.30$, bootstrapped $p = 0.257$; see **Table 3**). Thus, it is not the case that college-exposed participants' ability beliefs are predictive of gender gaps across academia simply because they are derived from prior knowledge of such gaps.

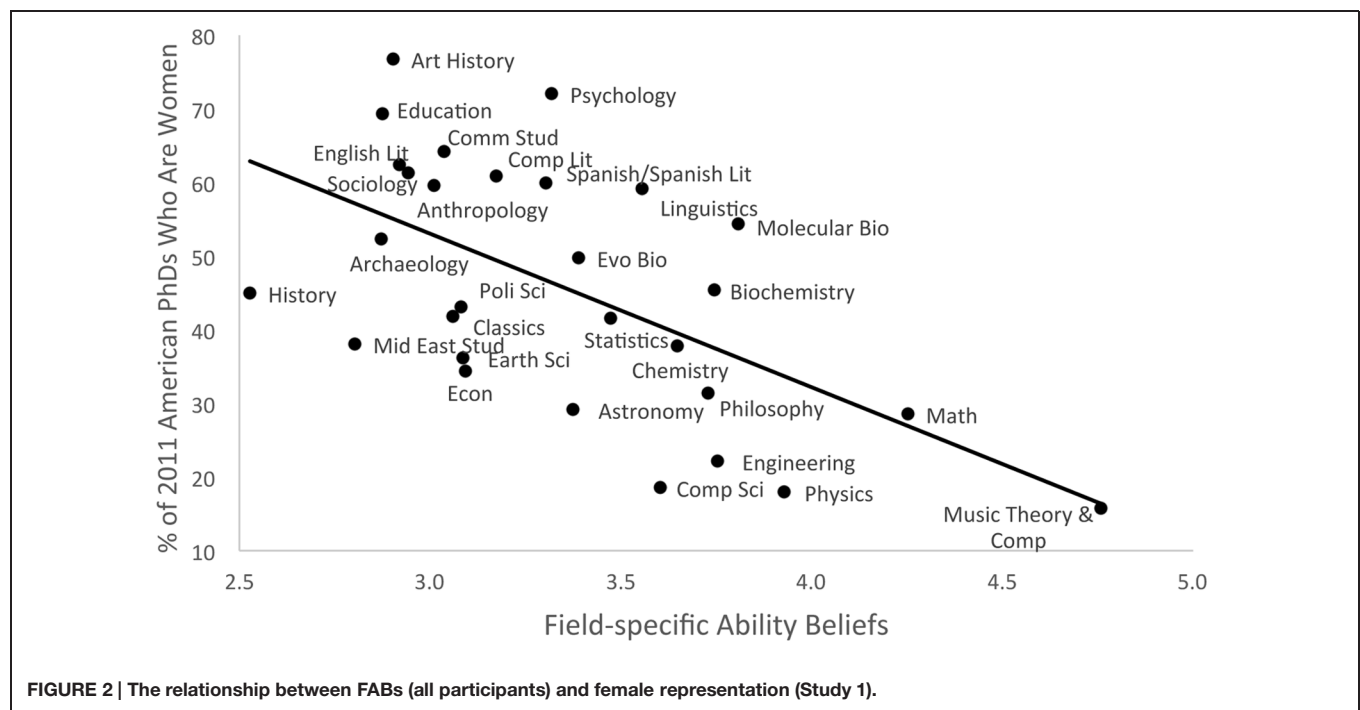


TABLE 3 | Regressions predicting female representation using field-specific ability beliefs and estimates of female representation of participants with college experience (CE; Upper) and with no college experience (NCE; Lower), Study 1.

Predictor	β	p	R^2	F	p
			0.686	15.87	<0.001
STEM indicator	0.07	0.686			
Estimate of female representation (CE)	0.58	0.001			
Field-specific ability beliefs (CE)	-0.41	0.043			
			0.55	10.37	<0.001
STEM indicator	-0.22	0.340			
Estimate of female representation (NCE)	0.44	0.023			
Field-specific ability beliefs (NCE)	-0.30	0.257			

$n = 29$, $df (4,24)$.

We considered one final alternative interpretation, which applies particularly to the findings obtained with college-exposed individuals. Perhaps College Exposure FAB scores emphasize brilliance for fields where there are few women just because (1) men may be more likely than women to believe that brilliance is required for success, and (2) more men in the current sample may have taken college classes in disciplines where women are typically underrepresented. In other words, disciplines with lower female representation may have higher College Exposure FAB scores for the simple reason that male participants' brilliance-focused ability beliefs are overrepresented in

our sample for these disciplines. Consistent with this possibility, college-exposed men's scores ($M = 3.56$, $SD = 0.55$) were indeed higher than college-exposed women's scores ($M = 3.18$, $SD = 0.64$), $t(28) = 4.02$, $p < 0.001$, suggesting that men placed more emphasis on raw ability. In addition, our sample contained proportionately more college-exposed men in fields with lower female representation at the Ph.D. level, $r(27) = -0.66$, $p < 0.001$. To test whether these differences could explain our main result, we calculated a gender-balanced FAB score for each field by computing the average scores for men and women separately within fields, and then averaging these two gender-specific scores. This measure adjusts for the differential representation of college-exposed males and females across fields, giving the two groups an equal say in determining the FAB score for each field. If the current alternative explanation were correct, this gender-balanced score should no longer be predictive of female representation. However, when we entered the gender-balanced FAB score in place of the original FAB score in the regression including both STEM status and estimated female representation, it still predicted female representation ($\beta = -0.35$, bootstrapped $p = 0.069$). Thus, the main results described above were not merely a byproduct of men's brilliance-oriented beliefs inflating the College Exposure FAB scores of fields with fewer women.

In sum, the results of Study 1 lend clear support to the predictions we derived from the FAB model: women are less likely to be represented in fields believed to require stable, innate ability. Furthermore, as predicted, the field-specific beliefs of people with college experience in our fields were predictive of female representation at a more detailed level than were the beliefs of those without college experience. To speculate, perhaps initially people hold a global belief that disciplines in the STEM family require innate skill; as a result, the predictive power of these

initial, inchoate ability beliefs is mostly captured by the STEM vs. non-STEM distinction. It is only after exposure to the particularities of the fields and the beliefs of their practitioners that FABs take on independent predictive power in relation to female representation.

Study 2 provides an opportunity to replicate the above findings, and to further explore how gender breakdowns are related to field-differentiated beliefs about the types of skills and work that are required. Two predictions are central to Study 2. First, we expect that the FABs of participants with college experience will predict unique variance in female representation, above and beyond their beliefs about the role of mathematical or verbal skills. Second, we predict that participants' assumptions about how much solitary and competitive work is required by individual fields will mediate the relationship between FABs and female representation.

Study 2

Method

Participants

Participants included 302 individuals recruited via Amazon's MTurk, using the same inclusion criteria as in Study 1. Participants were compensated \$0.95 for survey completion. Data were excluded from an additional 53 individuals who met one or more of the exclusion criteria used in Study 1: (1) failing to complete the survey, (2) answering an attention check question incorrectly, (3) having an IP address suggesting residence outside the U.S., and/or (4) having IP addresses indicating completion of similar studies (including our Study 1) in the past.

Materials and Procedure

As in Study 1, three versions of the survey were created, each of which contained the same subsets of 10 of the 30 fields under investigation. Approximately equal numbers of subjects participated in the three versions, and assignment was random (Version 1, $n = 101$; Version 2, $n = 103$; Version 3, $n = 98$). Surveys included the same four FAB items as in Study 1, along with two additional, broader questions on this topic (Table 2). These questions were added to further assess FABs, with the goal of using more accessible language while still providing a sensitive index of participants' beliefs about innate ability vs. effort. Two items were also included to address people's beliefs about the extent to which verbal and mathematical skills are required, and two final items were included to assess beliefs about whether competition and solitary work are important for success in a field (Table 2).

As in Study 1, items were presented individually in random order with all 10 fields listed beneath each item. Participants again indicated their agreement with the statement as it applied to each of the 10 fields using a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree, with eight as an option to indicate "don't know"). Two attention-check questions were also included. The survey then ended with questions assessing high school, college, and graduate level exposure to each of the 10 fields, along

with several demographic questions. (These questions were all identical to those in Study 1).

We calculated FAB scores by averaging scores across the six items, and then averaging within fields to create field-level scores. Three separate FAB scores were calculated reflecting (1) All Participants' FAB (using data from all participants except those with graduate level experience in the field), (2) College Exposure FAB (using data from participants who had taken college, but not graduate level, courses in the field), and (3) No College Exposure FAB (using data from participants who had taken neither college nor graduate courses in the field). Scores for the six ability beliefs questions had high internal reliability (for all participants, $\alpha = 0.89$; for College Exposure, $\alpha = 0.93$; for No College Exposure, $\alpha = 0.87$). Deletion of the last two items added for Study 2 did not improve scale reliability, indicating it was appropriate to include them as part of the FAB scale.

Results and Discussion

To explore whether Study 2 replicated the key finding that FABs predict female representation, we again examined correlations between FAB and percentage of female Ph.D. recipients. As before, fields with fewer than 10 participants reporting either college or no college exposure were removed. This resulted in 27 fields being retained for analysis. (Middle Eastern studies, neuroscience, and archeology were removed because they had College Exposure n s of 8, 4, and 3, respectively.) Replicating findings from Study 1, FAB scores were negatively associated with female representation when examining belief scores of all participants, $r(25) = -0.63$, $p < 0.001$, as well as when examining College Exposure scores, $r(25) = -0.65$, $p < 0.001$, and No College Exposure scores, $r(25) = -0.54$, $p = 0.004$. Although College Exposure scores were more strongly associated with female representation than No College Exposure scores, this difference was not significant according to Steiger's z score comparison (Lee and Preacher, 2013), $z = 1.13$, $p = 0.258$. However, again replicating Study 1, we found that the FABs of participants with college experience predicted female representation even when a STEM indicator variable was added to the regression model as a competitor ($\beta = -0.37$; bootstrapped $p = 0.048$); in contrast, the beliefs of those without college experience were not uniquely predictive when the STEM indicator was added ($\beta = -0.15$; bootstrapped $p = 0.48$). Thus, beliefs held by college-exposed individuals again predicted female representation better than those of non-college-exposed individuals.

We next examined the relationship between female representation and the extent to which a field is perceived as demanding verbal and mathematical skills. Beliefs about the need for verbal skills were positively associated with female representation: beliefs of all participants, $r(25) = 0.63$, $p < 0.001$; of participants with college exposure, $r(25) = 0.63$, $p = 0.001$; of participants with no college exposure, $r(25) = 0.65$, $p < 0.001$. Beliefs about the need for mathematical skills were negatively associated with female representation: beliefs of all participants, $r(25) = -0.64$, $p < 0.001$; of participants with college exposure, $r(25) = -0.60$, $p = 0.001$; of participants with no college exposure $r(25) = -0.64$, $p < 0.001$.

We then tested our prediction that FABs of individuals with college exposure would predict female representation independently from beliefs about the role of mathematical and verbal skills. If so, this would strengthen the claim that FABs tap into something distinct from people's beliefs about which fields require mathematical aptitude. To assess this prediction, we added perceptions of the need for verbal and mathematical skill as variables in the two regressions predicting female representation. For the regression testing beliefs of those *with* college experience (Table 4), FABs were uniquely predictive of women's representation, above and beyond STEM status and beliefs about the importance of mathematical and verbal skill, $\beta = -0.39$, bootstrapped $p = 0.085$, although this coefficient was only significant at the $\alpha = 0.10$ level. In contrast, beliefs about the importance of verbal and mathematical ability did not independently predict female representation in this model, $ps > 0.489$. For the regression testing the beliefs of those *without* college experience, no factor was significantly predictive of female representation, $ps > 0.454$ (Table 4).

As in Study 1, we also calculated a gender-balanced FAB score to examine the possibility that differences in male and female participants' ability beliefs and college experience were driving the effects observed for college-exposed participants. (To reiterate, the possibility being tested here is that College Exposure FAB scores in fields with fewer women are inflated simply because men may have ability beliefs that are more brilliance-oriented and may also be overrepresented in the college-exposure sample for these fields.) Again, the proportion of college-exposed male participants within each field was negatively related to female representation at the Ph.D. level, $r(25) = -0.41$, $p = 0.03$, indicating that college-exposed male participants were more numerous in fields with lower female representation. In this sample, however, college-exposed men's FAB scores ($M = 3.70$, $SD = 0.71$) were actually lower than college-exposed women's

scores ($M = 3.85$, $SD = 0.60$), though not significantly so, $t(26) = 1.69$, $p = 0.103$. Thus, this alternative explanation is unlikely: college-exposed male participants, though more numerous in fields with fewer women at the Ph.D. level, did *not* differ from college-exposed female participants in their ability beliefs. Nevertheless, we entered a gender-balanced FAB score in place of the original FAB score in the regression model that also included a STEM indicator, beliefs about mathematical ability, and beliefs about verbal ability. As before, ability beliefs were the sole predictor of female representation ($\beta = -0.44$, bootstrapped $p = 0.059$). These results strengthen the main claim that ability beliefs are predictive of female representation, above and beyond beliefs about mathematical and verbal skills.

Finally, we tested the prediction that beliefs about solo work and competitiveness would mediate the relationship between FABs and female representation. Consistent with our argument, a bootstrapped (1,000 replications) product-of-coefficients mediation analysis performed with the PROCESS procedure in SPSS 22 (Hayes, 2013) revealed that the relationship between college-exposed participants' ability beliefs about a discipline and the proportion of female Ph.D.'s in that discipline was significantly mediated by these participants' ideas about the amount of solo work and the level of competitiveness required by the discipline, $ab = -13.56$ (-26.74 , -2.91). Similar results were obtained when examining beliefs of non-college-exposed participants, $ab = -13.61$ (-24.65 , -5.94)⁵. (For full results of the mediation models, see Figures 3 and 4.) Results are thus consistent with the idea that FABs may influence women's participation in a field in part by influencing their beliefs about what it is like to be a member of that field—in particular, whether one works by oneself or with others, and whether success rests more on competition with colleagues rather than cooperation. Interestingly, this result was observed even within the group who had not had college exposure to the field, which may be because inferences about the nature of the work demanded by various fields are easily drawn from one's ability beliefs about these fields, no matter how much first-hand experience one has with them.

General Discussion

Women are underrepresented in many STEM fields, but the pattern of gender distribution is complex, and a substantial amount of variation also exists in non-STEM fields. An important aim of the current studies was to provide an account for the wide variability in female representation across the entire academic spectrum. We maintain that the FAB hypothesis provides such an account. This hypothesis predicts that women will be underrepresented in fields believed to emphasize brilliance and inherent ability as the key to success; this is because women are often stereotyped as lacking the same sort of innate intelligence as men, and thus women will be discouraged from participating in fields to the extent that these fields are perceived as requiring this type

TABLE 4 | Regressions predicting female representation using field-specific ability beliefs and beliefs about the importance of verbal and mathematical skill of participants with college experience (CE; Upper) and with no college experience (NCE; Lower), Study 2.

Predictor			R^2	F	p
	β	p	0.52	6.05	0.002
STEM indicator	-0.11	0.747			
Field-specific ability beliefs (CE)	-0.39	0.085			
Verbal skill beliefs (CE)	0.26	0.489			
Mathematical skill beliefs (CE)	-0.06	0.820			
	β	p	0.49	5.29	0.004
STEM indicator	-0.21	0.524			
Field-specific ability Beliefs (NCE)	-0.14	0.516			
Verbal skill beliefs (NCE)	0.17	0.614			
Mathematical skill beliefs (NCE)	-0.26	0.454			

$n = 27$, $df (4,22)$.

⁵The indirect paths were again significant even when adjusting for beliefs about mathematical skills, both for college-exposed participants, $ab = -9.21$ (-22.11 , -1.73) and for non-college-exposed participants, $ab = -9.54$ (-19.77 , -2.89).

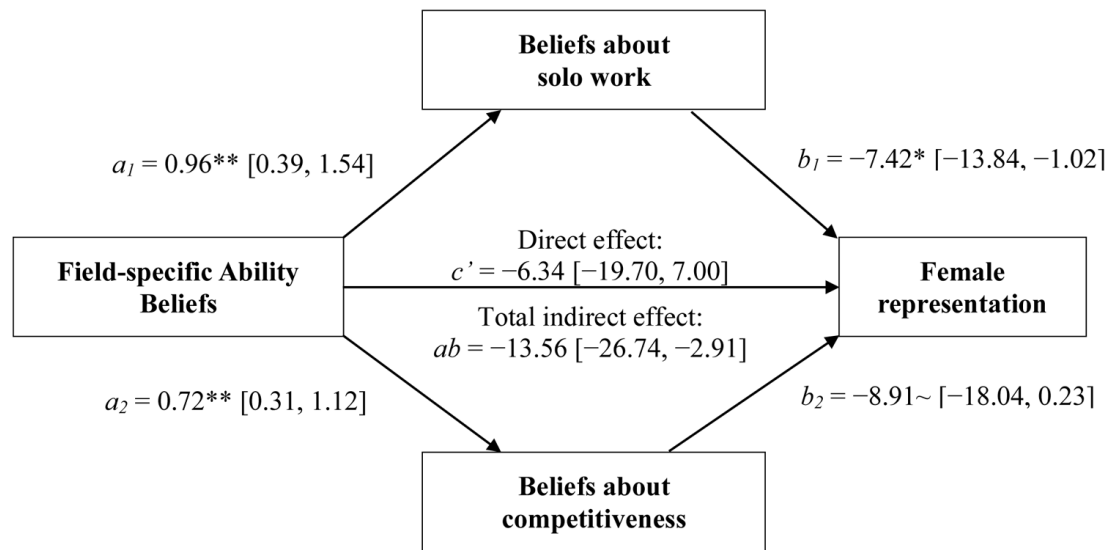


FIGURE 3 | The indirect pathways linking college-exposed participants' FABs with women's representation via participants' beliefs about the amount of solo work (top) and the level of competitiveness (bottom) required by a field (Study 2). $\sim p < 0.10$; $*p < 0.05$; $**p < 0.01$.

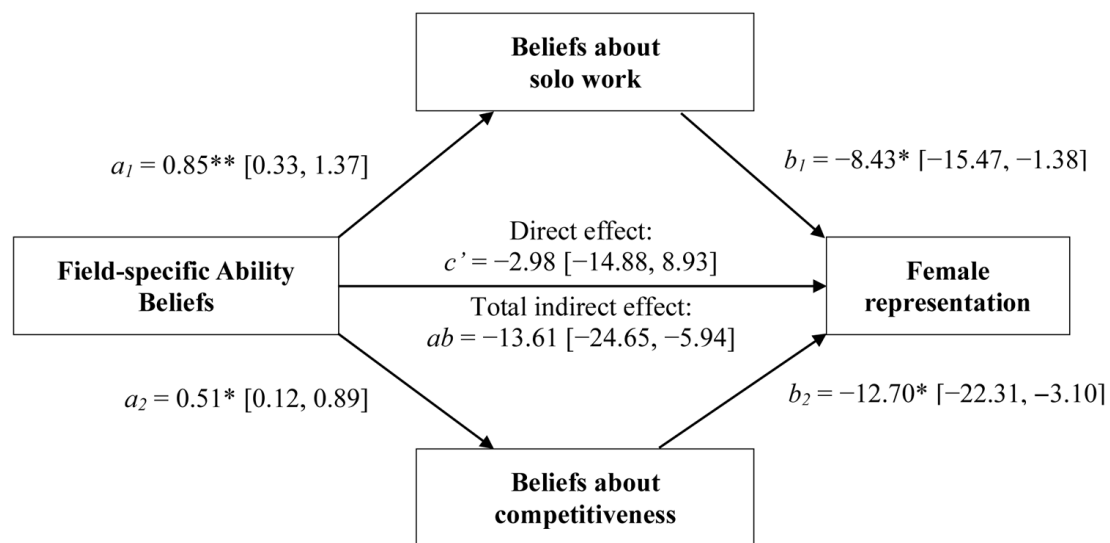


FIGURE 4 | The indirect pathways linking non-college-exposed participants' FABs with women's representation via participants' beliefs about the amount of solo work (top) and the level of competitiveness (bottom) required by a field (Study 2). $*p < 0.05$; $**p < 0.01$.

of intelligence. Prior research has provided support for the FAB hypothesis within higher academia (Leslie and Cimpian et al., 2015). The current studies extended the focus to an examination of beliefs held by individuals outside academia. The results of our two studies are consistent with the FABs hypothesis: fewer women are involved in fields that laypeople believe to require raw intellectual ability.

Several additional findings from the present studies are worth highlighting. The ability beliefs of individuals who had college-level exposure to the fields in question predicted female representation even when controlling for whether a field was in

STEM or not, indicating that college may provide a unique context for refinement and elaboration of beliefs about what fields require for success. Results also suggested that the ability beliefs of participants with college experience are not simply a byproduct of participants' inferring these beliefs based on their prior knowledge of female representation (Study 1). Further, college-exposed participants' ability beliefs capture something beyond perceptions of specific types of skills required for success, as FABs of college-exposed individuals did not reduce to beliefs about which fields require mathematical and/or verbal skills (Study 2).

Notably, these findings have important consequences for potential interventions to improve diversity, both in terms of timing and in terms of content. College may be a pivotal experience during which people's FABs become entrenched, and start to conform to those of their instructors. This highlights the crucial role that college educators play in communicating these maladaptive beliefs—but also suggests that they may be able to play an active role in changing the relevant messages. In particular, our data suggest that instructors who want to promote diversity might aim to minimize discussion of innate talent, regardless of the domain of skills with which it is associated, and instead highlight the importance of effort, practice, and persistence to success in a field. Prior work on individuals' achievement beliefs suggests that such growth-oriented messages can be relayed in a range of ways: by choice of adjectives (in particular by avoiding words like “brilliant,” “genius,” etc.; , Mueller and Dweck, 1998; Cimpian et al., 2007; Heyman, 2008), by focusing on what the person *has achieved* rather than on the person's *inherent traits* (Kamins and Dweck, 1999), and by explicitly stating that dedication and effort are paramount (Mueller and Dweck, 1998; Kamins and Dweck, 1999; Blackwell et al., 2007). We expect that practices such as these would be easily implementable by college educators across many fields. It should be borne in mind that the messages that college educators send may not only affect the participation of the women in their classes, but also have more far-reaching impact. As their students—both men and women—may go on to become parents, caregivers, school teachers, etc., they may subtly communicate their own ability beliefs to future generations (e.g., through their own choice of adjectives; Cimpian et al., 2007). This in turn may influence even very young girls' engagement and educational choices (Cimpian et al., 2014; Leslie et al., 2015).

The current studies also suggest that beliefs about solo and competitive work may mediate the relationship between ability beliefs and female representation. It is possible that this result reflects a process by which ability beliefs influence perceptions of what it is like to work in certain fields, which in turn may influence the participation of women in these fields. Of course, we acknowledge this is not the only possible pathway here; our mediation analyses were designed to test an a priori hypothesis regarding how ability beliefs relate to representation, but they cannot determine directionality. Similarly, causality regarding ability beliefs and female representation cannot be claimed from the current studies due to the correlational nature of the data. However, our theoretical model posits that ability beliefs do drive women's career and educational choices, and recent experimental manipulations in our lab have provided evidence consistent with this causal claim (e.g., Cimpian et al., 2014). For instance,

simply describing a novel educational or professional opportunity as requiring raw talent (vs. dedication) was sufficient to lower women's—and even young girls'—motivation to pursue it. Thus, there is some independent evidence suggesting that the relationship between ability beliefs and female representation is due to the causal influence of the ability beliefs.

Further investigating the precise pathways by which non-academics' ability beliefs influence participation is one important topic for future research. To begin, it is worth noting that young men and women often decide whether or not to pursue a field long before interacting with professors, graduate students, or any other active practitioners of that field (e.g., Watt and Eccles, 2008). Indeed, many fields with disproportionately high representation of men at the Ph.D. level see gender disparities in interest as early as elementary school (Lubinski and Benbow, 2006; Ceci and Williams, 2010; Cvencek et al., 2011). From our viewpoint, some of these early differences may be due to the ability beliefs of people outside of academia (teachers, parents, peers, etc.). For example, adults' FABs, in combination with the stereotype that females are less likely than males to be brilliant, could lead to small differences in the extent to which adults encourage girls' and boys' interest in fields believed to require this intellectual trait, the extent to which they provide boys and girls with opportunities to develop their skills in these fields, the extent to which they dwell on boys' and girls' achievements in these fields, and so on. Adults are also likely to convey their FABs to the children themselves. Once absorbed, these beliefs might make it more difficult for girls to consider careers in fields believed to require brilliance (again, since the ambient stereotypes portray them as being unsuited for these fields). As well, children might communicate these FABs to their peers, either via explicit statements or more subtly—say, by reacting with surprise to behaviors that are inconsistent with these beliefs. As a result of these multiple parallel processes, young women may be less likely to be interested in “brilliance-required” fields, and those who do pursue them may be less likely to persist and achieve at the same levels as men.

In summary, we have provided support for the FAB hypothesis, demonstrating that women tend to be underrepresented in fields believed to require innate intellectual talent for success. Our data also open up possibilities for future research on the pathways by which ability beliefs influence women's participation. Finally, these studies point to possibilities for effective interventions. If the practitioners of fields with gender gaps made a concerted effort to highlight the role of sustained, long-term effort in achievement, the gender gaps in these fields may correspondingly be diminished.

References

- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends Cogn. Sci.* 6, 248–254. doi: 10.1016/S1364-6613(02)01904-6
- Bennett, M. (1996). Men's and women's self-estimates of intelligence. *J. Soc. Psychol.* 136, 411–412. doi: 10.1080/0022454.1996.9714021
- Bennett, M. (1997). Self-estimates of ability in men and women. *J. Soc. Psychol.* 137, 540–541. doi: 10.1080/00224549709595475
- Bennett, M. (2000). Self-estimates and population estimates of ability in men and women. *Aust. J. Psychol.* 52, 23–28. doi: 10.1080/00049530008255363
- Berinsky, A. J., Huber, G. A., and Linz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis* 20, 351–368. doi: 10.1093/pan/mpr057
- Blackwell, L. S., Trzesniewski, K. H., and Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Dev.* 78, 246–263. doi: 10.1111/j.1467-8624.2007.00995.x

- Ceci, S. J., Ginther, D., Kahn, S., and Williams, W. M. (2014). Women in academic science: a changing landscape. *Psychol. Sci. Public Interest*. 15, 75–141. doi: 10.1177/1529100614541236
- Ceci, S. J., and Williams, W. M. (2007). *Why aren't More Women in Science? Top Researchers Debate the Evidence*. Washington, DC: American Psychological Association.
- Ceci, S. J., and Williams, W. M. (2010). Sex differences in math-intensive fields. *Curr. Dir. Psychol. Sci.* 19, 275–279. doi: 10.1177/0963721410383241
- Ceci, S. J., Williams, W. M., and Barnett, S. M. (2009). Women's underrepresentation in science: sociocultural and biological considerations. *Psychol. Bull.* 135, 218–261. doi: 10.1037/a0014412
- Cimpian, A., Arce, H. C., Markman, E. M., and Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychol. Sci.* 18, 314–316. doi: 10.1111/j.1467-9280.2007.01896.x
- Cimpian, A., Bian, L., and Leslie, S.-J. (2014). Experimental and developmental evidence for the field-specific ability beliefs hypothesis. *Paper presented at the Fifteenth Annual Meeting of the Society for Personality and Social Psychology*, Austin, TX.
- Cvencek, D., Meltzoff, A., and Greenwald, A. (2011). Math-gender stereotypes in elementary school children. *Child Dev.* 82, 766–779. doi: 10.1111/j.1467-8624.2010.01529.x
- Diekmann, A. B., Brown, E. R., Johnston, A. M., and Clark, E. K. (2010). Seeking congruity between roles and goals: a new look at why women opt out of STEM careers. *Psychol. Sci.* 21, 1051–1057. doi: 10.1177/0956797610377342
- Dweck, C. S. (2000). *Self-theories: Their Role in Motivation, Personality, and Development*. Philadelphia, PA: Psychology Press.
- Dweck, C. S. (2006). *Mindset: The New Psychology of Success*. New York: Random House.
- Ferriman, K., Lubinski, D., and Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: developmental changes and sex differences during emerging adulthood and parenthood. *J. Pers. Soc. Psychol.* 97, 517–532. doi: 10.1037/a0016030
- Fredericks, J. A., and Eccles, J. E. (2002). Children's competence and value beliefs from childhood through adolescence. *Dev. Psychol.* 38, 519–533. doi: 10.1037/0012-1649.38.4.519
- Furnham, A., Crawshaw, J., and Rawles, R. (2006). Sex differences in self-estimates on two validated IQ test subscale scores. *J. Appl. Psychol.* 36, 417–440. doi: 10.1111/j.0021-9029.2006.00013.x
- Gupta, N. G., Poulsen, A., and Villeval, M. C. (2013). Gender matching and competitiveness: experimental evidence. *Econ. Inq.* 51, 816–835. doi: 10.1111/j.1465-7295.2011.00378.x
- Hayes, A. F. (2013). *Introduction to Mediation, Moderation and Conditional Process Analysis*. New York: Guilford Press.
- Hedges, L. V., and Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science* 269, 41–45. doi: 10.1126/science.7604277
- Herbert, J., and Stipek, D. T. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Appl. Dev. Psychol.* 26, 276–295. doi: 10.1016/j.appdev.2005.02.007
- Heyman, G. D. (2008). Talking about success: implications for achievement motivation. *J. Appl. Dev. Psychol.* 29, 361–370. doi: 10.1016/j.appdev.2008.06.003
- Hill, C., Corbett, C., and St. Rose, A. (2010). *Why so few? Women in Science, Technology, Engineering, and Mathematics*. Washington, DC: American Association of University Women.
- Kamins, M. L., and Dweck, C. S. (1999). Person versus process praise and criticism: implications for contingent self-worth and coping. *Dev. Psychol.* 35, 835–847. doi: 10.1037/0012-1649.35.3.835
- Kirkcaldy, B., Noack, P., Furnham, A., and Siefen, G. (2007). Parental estimates of their own and their children's intelligence. *Eur. Psychol.* 12, 173–180. doi: 10.1027/1016-9040.12.3.173
- Lecklider, A. (2013). *Inventing the Egghead*. Philadelphia, PA: University of Pennsylvania Press. doi: 10.9783/9780812207811
- Lee, I. A., and Preacher, K. J. (2013). *Calculation for the Test of the Difference Between Two Dependent Correlations with one Variable in Common [Computer software]*. Available at: <http://quantpsy.org>
- Leslie, S.-J., Cimpian, A., Meyer, M., and Freeland, E. (2015). Women are underrepresented in disciplines that emphasize brilliance as the key to success. *Science* 347, 262–265.
- Leslie, S.-J., Saunders, K., and Rhodes, M. (2015). “Let's be scientists!": Hearing generic language about scientists lowers preschool-aged girls' motivation and engagement with a science task,” in *Proceedings of the Symposium Conducted at the Meeting of the Society for Personality and Social Psychology: Bringing Thought Experiments to Life: Perspectives on the Essence of Identity from Psychology and Philosophy*, eds L. Heiphetz and A. Waytz (Chairs), Long Beach, CA.
- Lippa, R. A. (1998). Gender-related individual differences and the structure of vocational interests: the importance of the people-things dimension. *J. Pers. Soc. Psychol.* 74, 996–1009. doi: 10.1037/0022-3514.74.4.996
- Lippa, R. A., Preston, K., and Penner, J. (2014). Women's representation in 60 occupations from 1972 to 2010: more women in high-status jobs, few women in things-oriented jobs. *PLoS ONE* 9:e95960. doi: 10.1371/journal.pone.0095960
- Lubinski, D., and Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: uncovering antecedents for math science expertise. *Perspect. Psychol. Sci.* 1, 316–345. doi: 10.1111/j.1745-6916.2006.00019.x
- Mueller, C. M., and Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *J. Pers. Soc. Psychol.* 75, 33–52. doi: 10.1037/0022-3515.75.1.33
- National Science Foundation [NSF]. (2011). *Survey of Earned Doctorates*. Available at: <http://www.nsf.gov/statistics/srvydoctorates/> (2011).
- Niederle, M., and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Q. Econ.* 122, 1067–1101. doi: 10.1162/qjec.122.3.1067
- Paolacci, J., and Chandler, J. (2014). Inside the turk: understanding mechanical turk as a participant pool. *Curr. Dir. Psychol. Sci.* 23, 184–188. doi: 10.1177/0963721414531598
- Rammstedt, B., and Rammsayer, T. H. (2002). Self-estimated intelligence: gender differences, relationship to psychometric intelligence and moderating effects of level of education. *Eur. Psychol.* 7, 275–284. doi: 10.1027/1016-9040.7.4.275
- Shenk, J. W. (2014). The end of 'genius.' *The New York Times*. Available at: <http://www.nytimes.com>
- Stephens-Davidowitz, S. (2014). *Google, Tell Me: Is My Son a Genius?* Available at: http://www.nytimes.com/2014/01/19/opinion/sunday/google-tell-me-is-my-son-a-genius.html?_r=0 [accessed October 16, 2014].
- Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Educ. Stud. Math.* 50, 40–62.
- Upson, S., and Friedman, L. F. (2012). Where are all the female geniuses? *Sci. Am. Mind* 23, 63–65. doi: 10.1038/scientificamericanmind1112-63
- Watt, H. M. G., and Eccles, J. S. (eds). (2008). *Gender and Occupational Outcomes: Longitudinal Assessments of Individual, Social, and Cultural Influences*. Washington, DC: American Psychological Association.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Meyer, Cimpian and Leslie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Women in Academic Science: A Changing Landscape

Authors: Stephen J. Ceci, Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams

[Volume 15, Issue 3](#)

<https://doi.org/10.1177/1529100614541236>

<https://journals.sagepub.com/doi/abs/10.1177/1529100614541236>

Abstract

Much has been written in the past two decades about women in academic science careers, but this literature is contradictory. Many analyses have revealed a level playing field, with men and women faring equally, whereas other analyses have suggested numerous areas in which the playing field is not level. The only widely-agreed-upon conclusion is that women are underrepresented in college majors, graduate school programs, and the professoriate in those fields that are the most mathematically intensive, such as geoscience, engineering, economics, mathematics/computer science, and the physical sciences. In other scientific fields (psychology, life science, social science), women are found in much higher percentages.

In this monograph, we undertake extensive life-course analyses comparing the trajectories of women and men in math-intensive fields with those of their counterparts in non-math-intensive fields in which women are close to parity with or even exceed the number of men. We begin by examining early-childhood differences in spatial processing and follow this through quantitative performance in middle childhood and adolescence, including high school coursework. We then focus on the transition of the sexes from high school to college major, then to graduate school, and, finally, to careers in academic science.

The results of our myriad analyses reveal that early sex differences in spatial and mathematical reasoning need not stem from biological bases, that the gap between average female and male math ability is narrowing (suggesting strong environmental influences), and that sex differences in math ability at the right tail show variation over time and across nationalities, ethnicities, and other factors, indicating that the ratio of males to females at the right tail can and does change. We find that gender differences in attitudes toward and expectations about math careers and ability (controlling for actual ability) are evident by kindergarten and increase thereafter, leading to lower female propensities to major in math-intensive subjects in college but higher female propensities to major in non-math-intensive sciences, with overall science, technology, engineering, and mathematics (STEM) majors at 50% female for more than a decade. Post-college, although men with majors in math-intensive subjects have historically chosen and completed PhDs in these fields more often than women, the gap has recently narrowed by two thirds; among non-math-intensive STEM majors, women are more likely than men to go into health and other people-related occupations instead of pursuing PhDs.

Importantly, of those who obtain doctorates in math-intensive fields, men and women entering the professoriate have equivalent access to tenure-track academic jobs in science, and they persist and are remunerated at comparable rates—with some caveats that we discuss. The transition from graduate programs to assistant professorships shows more pipeline leakage in the fields in which women are already very prevalent (psychology, life science, social science) than in the math-intensive fields in which they are underrepresented but in which the number of females holding assistant professorships is at least commensurate with (if not greater than) that of males. That is, invitations to interview for tenure-track positions in math-intensive fields—as well as actual employment offers—reveal that female PhD applicants fare at least as well as their male counterparts in math-intensive fields.

Along these same lines, our analyses reveal that manuscript reviewing and grant funding are gender neutral: Male and female authors and principal investigators are equally likely to have their manuscripts accepted by journal editors and their grants funded, with only very occasional exceptions. There are no compelling sex differences in hours worked or average citations per publication, but there is an overall male advantage in productivity. We attempt to reconcile these results amid the disparate claims made regarding their causes, examining sex differences in citations, hours worked, and interests.

We conclude by suggesting that although in the past, gender discrimination was an important cause of women's underrepresentation in scientific academic careers, this claim has continued to be invoked after it has ceased being a valid cause of women's underrepresentation in math-intensive fields. Consequently, current barriers to women's full participation in mathematically intensive academic science fields are rooted in pre-college factors and the subsequent likelihood of majoring in these fields, and future research should focus on these barriers rather than misdirecting attention toward historical barriers that no longer account for women's underrepresentation in academic science.