DISCUSSION PAPER SERIES

IZA DP No. 5537

Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment

Natalia Zinovyeva
Manuel Bagues

February 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment

Natalia Zinovyeva
FEDEA, Madrid

Manuel Bagues
Universidad Carlos III de Madrid
and IZA

Discussion Paper No. 5537
February 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5537
February 2011

ABSTRACT

Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment*

Several countries have recently introduced gender quotas in hiring and promotion committees at universities. This paper studies whether these policies increase the presence of women in top academic positions. The identification strategy exploits the random assignment mechanism in place between 2002 and 2006 in all academic disciplines in Spain to select the members of promotion committees. We find that a larger proportion of female evaluators increases the chances of success of female applicants to full professor positions. The magnitude of the effect is large: each additional woman on a committee composed of seven members increases the number of women promoted to full professor by 14%. Conversely, when committee members decide on promotions to associate professor positions, we do not observe any significant interaction between the gender of evaluators and the gender of candidates. If anything, in this case a larger share of female evaluators is associated with fewer successful female applicants. The evidence is consistent with the existence of ambivalent sexism.

JEL Classification: J71, J45

Keywords: academic promotion, gender discrimination, randomized natural experiment

Corresponding author:

Natalia Zinovyeva
IPP-CSIC
C/Albasanz, 26-28
28037 Madrid
Spain
E-mail: natalia.zinovyeva@cchs.csic.es

* We would like to thank Olympia Bover, Irma Clots-Figueras, Sara de la Rica, Juanjo Dolado, David Dorn, Berta Esteve-Volart, Luis Garicano, Marco Giarratana, Elena Martínez, Nic Morgan, Javier Ruiz-Castillo and participants in presentations at Universidad del País Vasco, CEMFI, IPP-CSIC, Universidad de Granada, SOLE-EALE, Universitat Pompeu Fabra, Universitat Autònoma de Barcelona, GSOM St. Petersburg and Università Tor Vergata for their useful comments. This paper was prepared while the second author was visiting the Bank of Spain and CEMFI, whose hospitality is gratefully acknowledged. 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

Women have historically been under-represented in top academic positions.[1] In the past, this under-representation was partly the result of the smaller number of women obtaining doctorates. Policies toward gender equality were focused on the so-called "equal opportunities approach". Underlying this approach was the pipeline theory, according to which women must move their way through a metaphorical pipeline to reach top-level jobs. Accordingly, policy was designed to encourage women's higher education on the understanding that providing women with the same human capital as men would enable them to reach the top positions they seemed otherwise unable to attain. Evidence supporting the pipeline theory, however, is disappointing. While the number of women undertaking PhD studies has increased steadily, the incidence of women in the upper echelons of the academic career ladder remains low, particularly among full professors. For instance, in Spain, the presence of women among PhD graduates has grown from 36% to 49% over the last twenty years. During the same period, the incidence of women among faculty has increased from 30% to 39% among associate professors, but only from 11% to 18% among full professors (Figure 1). The picture is qualitatively similar in the U.S. and the rest of Europe.[2]

It could be that women differ in some characteristics that divert them from advancement in their academic career. Some authors argue that family commitments make it more difficult for women to move up the academic career ladder beyond their early post-doctorate years (National Research Council 2007). Women's careers may also be hindered by the lack of role models among the upper echelons (Holmes and O'Connell 2007). Alternatively, women may face discrimination, either explicit or implicit, in

[1]The lack of women in top academic positions has been documented in Life Sciences (Ginther and Kahn 2009), in the Humanities (Ginther and Hayes 2003), in Economics (McDowell et al. 1999, Ginther and Kahn 2004, Blackaby et al. 2005) and in Medicine (Tesch et al. 1995). Two recent National Research Council reports (2007, 2010) analyze the low presence of women in more senior academic positions and review the related literature.

[2]In Europe, women account for 45% of PhD graduates, 36% of associate professors and a mere 18% of full professors (European Commission 2009). In the U.S., excluding the Humanities, the incidence of women among new PhDs was around 40%; figures are 34% and 19% for associate professors and full professors respectively (National Science Foundation 2009).

promotions to top positions by the (mostly male) committees granting promotion.[3] To prevent gender discrimination, several countries, including Norway (1988), Finland (1995), Sweden (1999) and Spain (2007), have introduced a minimum share of women in hiring and promotion academic committees (European Commission 2008). In principle, if male evaluators are biased against women and female evaluators are more objective, the introduction of gender quotas in hiring and promotion committees should improve the chances of success of female applicants. However, almost no empirical evidence exists about the effectiveness of these policies.

A neat empirical analysis is usually hard to come by. One problem has been the lack of female evaluators. For instance, Wennerås and Wold (1997) study gender discrimination in applications to postdoctoral fellowships in Sweden but cannot address whether the gender of evaluators matters due to the paucity of women among peer reviewers (only five out of 55). Similarly, Combes et al. (2008) analyze the determinants for promotion in Economics departments in France, where only three out of seventy evaluators were female. Another important handicap has been the endogeneity of committee composition. In most situations, it cannot be ruled out that the gender composition of hiring committees is related to the relative quality of female and male candidates.

The available empirical evidence dealing with endogeneity provides results contradicting the view that gender quotas are effective. Broder (1993) examines the ratings of proposals to grants from the National Science Foundation (NSF). She finds that female reviewers rate female-authored NSF proposals lower than do their male colleagues. Following a similar identification strategy, Abrevaya and Hamermesh (2011) examine referee evaluations in a leading journal in Economics and do not find any effect from the interaction between the gender of referees and the gender of authors. In another setting, Bagues and Esteve-Volart (2010) analyze hiring for entry-level positions in the

[3]For instance, a report by a Spanish governmental organization, the Foundation for Science and Technology, claims that in "academia, promotion is based on a system [...] that benefits men more than women, since the barriers arise when mostly male committees evaluate female candidates and reject their promotion", "Mujer y Ciencia: La situación de las Mujeres Investigadoras en el Sistema Español de Ciencia y Tecnología", Fundación Española para la Ciencia y la Tecnología, 2005, p. 48.

Spanish Judiciary and find that female candidates are significantly less likely to be hired if they are (randomly) assigned to a committee with a relatively greater proportion of female evaluators. Booth and Leigh (2010) conduct an audit study in several female-dominated occupations in Australia and do not find any significant interaction between the gender of the applicant and the gender of the contact person in the hiring firm.

This body of evidence suggests that policy makers should not take female evaluators' alignment with female candidates as granted; however, none of these studies considers promotion to top positions. Given the lower number of women in top positions, and the persistence of the so-called *glass ceiling*, additional evidence is needed in order to understand the role of the gender of evaluators for promotion to higher level academic positions.

In this paper we address this issue using evidence from promotions in the Spanish university system between 2002 and 2006. During this period, all academic promotions were decided through nation-wide competitions. Our setup has three exceptional features. First, evaluations were performed for two types of positions: associate professor and full professor positions. Second, the system affected a large number of candidates, as well as evaluators from all academic disciplines. In total, approximately 35,000 candidacies were evaluated by 7,000 evaluators. Third, evaluators were selected out of a pool of eligible professors using a lottery. The existence of a system of random assignment of evaluators to committees allows us to consistently estimate the effect of the gender composition of committees. To our knowledge, this is the first study that exploits a randomized natural experiment in order to analyze the determinants of promotion to top positions.

We find that the gender composition of committees strongly affects the chances of success of candidates applying to full professor positions. In quantitative terms, for a committee with seven members, each additional female evaluator increases the chances of success of female applicants by 14%. When evaluators decide on promotions to associate professor positions, we fail to observe any significant interaction between the gender of evaluators and the gender of candidates.[4] If anything, a larger presence of women in the committee may decrease the number of female candidates promoted to associate professor.

To investigate which committees are gender-biased in exams to full professor, we have collected information on the research productivity of candidates from the ISI Web of Science. Specifically, we observe the number of publications and the number of citations received by candidates. As long as these measures capture all potential gender differences in quality, taking them into account allows us to estimate the source of discrimination. We find that, conditional on the research production of candidates, female applicants to full professor positions have lower probability of success relative to male candidates when assigned to an all-male committee. In committees including at least one woman, we do not observe significant gender differences in success rates.

We explore several potential explanations consistent with our evidence. We find that this gender bias exists in small disciplines, but not in large disciplines. This goes against information-based explanations of the gender bias: in large disciplines, people are less likely to know each other prior to the exam, and thus information asymmetries should be larger. Our findings are also not consistent with female candidates being discriminated against on the basis of age, or with the existence of gendered networks. Instead, the evidence in this paper might reflect the existence of ambivalent sexism, arising when men's attitudes toward female candidates depend on the position at stake. Specifically, male evaluators might experience sexist antipathy towards female candidates applying to top academic positions, but not towards female applicants to lower-level positions.

The rest of the paper is organized as follows. Section 2 explains the institutional background. Section 3 describes the data. In Section 4 we analyze whether the gender composition of committees affects promotion, and in Section 5 we explore potential explanations for our results. Section 6 concludes.

[4]The magnitude of the effect is similar to the effect found in Bagues and Esteve-Volart (2010), but it is not significant at standard levels. In what follows, we consider as standard a statistical significance level of 5%.

2 Institutional background

In Spain, approximately 88% of university professors work at public universities (Instituto Nacional de Estadística 2010). Before 2002, public universities had a large degree of autonomy regarding hiring and promotion. Some have criticized the large degree of inbreeding in the Spanish academe: 93% of the positions were assigned to candidates that attended the university offering the position (Cruz-Castro et al. 2006). In order to increase transparency and meritocracy in the evaluation procedure, in 2002, the government introduced a system of centralized competition known as *habilitación*.[5] *Habilitación* is relatively similar to promotion systems currently in place in France and Italy.[6] This system required candidates to associate and full professor positions to qualify in national competitions held at the discipline level.[7,8] Successful candidates could then apply for a position at a given university. In practice, the number of vacancies opened at the national level was very limited and the competition at the university level was almost absent. Being accredited was, in most cases, equivalent to being promoted.

The time line of examinations was as follows. First, the centralized competition was announced, and candidates were allowed twenty days to apply. Once the list of applicants was settled, committee members were selected by random draw from the list of eligible evaluators. This list included professors who were officially recognized to have a minimum research quality in the discipline.[9] Each committee was composed of seven

[5] Julio Iglesias de Ussel, vice-minister for Education and Universities, newspaper El País, November 5th, 2001.

[6] In France, professors in many disciplines are recruited through a centralized examination (*concours nationaux d'agrégation*). In Italy, the Moratti Law (2005) introduced a nation-wide qualification exam for candidates to university positions (*l'idoneità nazionale*).

[7] The position of *catedrático de universidad* at a Spanish university may be considered equivalent to the position of full professor in a U.S. university. The category of *profesores titulares de universidad* would be equivalent to associate professor; in Spain, the position of associate professor always carries tenure.

[8] In total, there are nearly two hundred legally defined academic disciplines, each corresponding to a certain field of knowledge. These disciplines were created in 1984 on the basis of "the homogeneity of its object of knowledge, a common historic tradition and the existence of a community of researchers" (R.D. 1988/84).

[9] The research quality requirement was based on the number of *sexenios* recognized to each professor. *Sexenios* 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. In particular, eligible assistant professors were required to hold at least one *sexenio*. Eligible full professors were required to hold at

members. For exams to full professor positions, all committee members were chosen from the list of eligible full professors. For exams to associate professor positions, three committee members were selected from the set of eligible full professors, and four were selected from the set of eligible associate professors.[10] The committee member with the longest tenure was appointed president, and the exam was held at the university where the president was based. Evaluators could only resign under a very restricted set of reasons, and their resignation had to be officially approved.[11]

Exams to full professor positions had two qualifying stages; first, every candidate presented her *résumé*, second, candidates presented a piece of their research work. Exams to associate professor also had an intermediate stage where candidates gave a lecture on a topic in their discipline, which was randomly chosen from a syllabus proposed by the candidate.

In 2006, the system of *habilitación* was replaced by a system known as *acreditación*, which is still in place. As in the system of *habilitación*, candidates are required to be accredited by a national evaluation committee. However, under the new system, committee members are selected from the pool of professors that volunteer for the task, and there is no limit to the number of candidates that may receive the accreditation. The 2007 Equality Law mandated gender parity in evaluation committees.

3 Data

We have collected data from all exams for associate and full professor positions that were held in Spain when the centralized system of examinations known as *habilitación* was in place (years 2002 through 2006). In total, 1016 exams took place over the period, around five per discipline. We restrict the sample in several ways. First, we

least two *sexenios*.

[10]Approximately 5% of eligible evaluators were at the Spanish National Research Council (CSIC). According to the rules, not more than one CSIC researcher was allowed to be selected as a member of the evaluation committee for a given exam. Similarly, not more than one emeritus professor was allowed to be selected as a member of a given evaluation committee. Whenever a second CSIC researcher or a second emeritus professor was drawn in the lottery, this draw was not considered.

[11]This happened very rarely. According to our own calculations using data for the year 2005, less than 3% of the rostered evaluators were replaced. They were substituted by a randomly selected evaluator.

exclude disciplines where the number of potential evaluators was not big enough to form a committee (55 exams).[12] We exclude exams where the population of potential evaluators did not include any women (55 exams), exams where all candidates were of the same gender (13 exams), and exams where the number of available positions was larger than or equal to the number of candidates (two Basque Philology exams). The final database includes 891 exams, of which 455 are exams to associate professor positions and 436 are exams to full professor positions.

Table 1 provides descriptive information on the characteristics of exams. On average, for full professor exams there were three available positions per exam. For associate professor exams, the figure was around five available positions. The number of positions per candidate was very similar in both types of competitions; around 0.12 positions per candidate. Most vacancies were filled: 98% in exams to full professor positions, and 96% in exams to associate professor positions. There are some differences across disciplinary areas in terms of the number of available positions per applicant. The fields with the lowest ratios are Mathematics and Physics. For these, on average there were 0.08 positions per candidate in full professor exams, and 0.09 in associate professor exams. The highest ratios can be found in Biology and Chemistry, with 0.13 and 0.15 positions per candidate in exams to full and associate professor positions respectively.

The upper panel of Table 2 provides information on the observable characteristics of eligible professors for evaluation committees. Note that, as explained in section 2, eligible professors are a selected sample of all professors. In order to be eligible as evaluators, professors were required to have a minimum level of research production. This minimum research level was satisfied by a relatively larger proportion of female than male professors. Around 80% of male and 84% of female full professors, and approximately 69% of male and 72% of female assistant professors qualified.[13]

In the sample, there are 21,944 eligible associate professors and 7,909 eligible full

[12] In theses cases, unfilled seats in the committee were filled with professors from related disciplines.

[13] Source: *Comisión Nacional Evaluadora de la Actividad Investigadora*, Memoria de los resultados de las evaluaciones realizadas de 1989 a 2005, 2005.

professors. We observe their gender, age, tenure and research production. Details about how these variables have been constructed are available in Appendix A. The share of women decreases higher up the career ladder. Women constitute 35% of associate professors but only 14% of full professors. The average eligible full professor is 53 years old; the average eligible associate professor is 45. Male full professors have longer tenure and tend to be older than female full professors, but among associate professors we observe that women are older and have slightly longer tenure.

Information on research output was collected from the Web of Science (ISI). Since ISI does not accurately capture research performance in the Humanities, we do not consider research productivity in this disciplinary area. The average number of publications – weighted by the number of co-authors – is equal to ten works for full professors and equal to six for associate professors. On average, full professors' publications have received eleven citations; associate professors' publications have received around ten citations. Since disciplines differ in their propensity to publish and to cite, we also provide information on the number of publications and the number of citations per publication normalized to have zero mean and unit standard deviation for professors within the same discipline and year. This normalization allows us to compare research production across disciplines, and gives us the research measure we use in our empirical analysis. Female eligible professors have slightly lower research output than male professors, as measured by both the normalized number of publications and the normalized number of citations per publication. Taking into account that the sample of eligible professors only includes professors whose research quality was above a certain quality threshold (which women were more likely to pass), the descriptive evidence suggests that female professors are more likely to have a minimum level of research quality but that (conditional on achieving this minimum) they tend to be slightly worse than their male counterparts.

Information on candidates is provided in the lower panel of Table 2. There are 13,224 applications to full professor positions, and 18,792 applications to associate professor positions. On average, candidates applied approximately twice during the period of study. The total number of candidates is 6,037 for full professor exams and 9,952 for associate professor exams. There are relatively fewer female applicants in exams for full professor positions. The share of women among candidates to full and associate professor positions is equal to 28% and 40% respectively. Not surprisingly, candidates to full professor positions tend to be older: 46 year old vs. 37 years old. There are some gender differences in terms of the age of applicants. Male candidates to full professor positions are slightly younger than female candidates, but this difference is not statistically significant. Male applicants to associate professor positions tend to be approximately one year older than their female counterparts. On average, applicants to associate professor positions have published five publications, and they have received around nine citations per publication. Applicants to full professor positions have accumulated a stronger research record: they have seven publications, and they have received eleven citations per publication. Both in exams to associate and full professor positions, male applicants have relatively more publications, but this difference is only statistically significant in full professor exams. Finally, there are no significant gender differences in terms of the number of received citations by female-authored and male-authored work.

Table 3 provides information about the degree of feminization of the different disciplinary areas. The least feminized disciplinary area is Engineering, where only 5% of full professors are women. The proportion of female full professors is slightly larger in Mathematics and Physics (8%), and Medicine (12%). The more feminized disciplinary areas are Social Sciences (16%), Biology and Chemistry (17%), and the Humanities (18%). The same pattern is observed if we examine the proportion of women among applicants to full professor positions, associate professors, and applicants to associate professor positions.

4 Empirical analysis

Our empirical analysis is structured as follows. First, we investigate whether the gender composition of academic committees affects applicants' chances of being promoted.

Then, using information on applicants' research productivity, we analyze which committees discriminate.

4.1 Does the gender composition of committees matter?

To analyze whether the gender composition of committees affects academic promotion, we compare the outcomes of candidates who applied to exams where the expected proportion of women in the committee (μ_e) was similar, but the realization of the random draw resulted in committees with a different gender composition (s_e). In other words, our identification strategy exploits the fact that, because of random sampling, $E[s_e|\cdot] - \mu_e = 0$.

Note that in exams to full professor positions, the expected proportion of women in the evaluation committee is essentially equal to the proportion of women in the pool of eligible full professors.[14] In exams to associate professor positions, three evaluators are drawn from the pool of eligible full professors, and four evaluators are drawn from the pool of eligible associate professors. Therefore, in this case, the expected proportion of women in the committee is a weighted average of the proportion of women in each of the two pools.

As an illustration of how the identification strategy works, let us consider the examinations to full professor positions in Economics between 2002 and 2006. During this period six examinations took place. The population of eligible evaluators included approximately seventy male and twelve female full professors. On average, evaluation committees were expected to be formed by six men and one woman. As a result of the random assignment mechanism, on three occasions there were six men and one woman on the committee, in one occasion all seven evaluators were male, and in two occasions, there were two women. Given that the pool of applicants was formed before the committee composition was decided, there is no reason to expect that the relative

[14] As explained in footnote 10, the random assignment of evaluators to committees was subject to a constraint: every committee could include at most one CSIC researcher and one emeritus professor. Therefore, in exams where the population of potential evaluators contained two or more researchers, or two or more emeritus professors, the expected proportion of women in the committee should be computed taking into account this constraint. This affects 300 of the 891 exams in the sample. The details on these calculations are in Appendix B.

quality of male and female candidates in each exam was related to random variations in the composition of evaluation committees.

The key assumption in our identification strategy is that the selection of committee members was really random. The selection was carried out by Ministry officials following a computerized random procedure certified by notary. Table 4 presents comparative information on the expected composition of committees and on the committees that were actually drawn by the lottery. The actual composition of committees is statistically similar to the expected composition in terms of all observable characteristics, which is consistent with a random assignment. In the case of full professor exams, we observe that the expected proportion of female evaluators and the actual proportion of women sitting on committees were both equal to 0.15, around 36% of committees were composed only by male evaluators, and only in 3% of the cases were there four or more female evaluators in the committee (columns 1 and 2). In exams to associate professor positions, the share of women was equal to 0.26, approximately 18% of committees did not include any female member, and 10% of exams had a majority of female evaluators (columns 3 and 4).

In our data, we observe the outcome of the random lottery but we do not observe who actually sat on the committee. There are two possible sources of variation. First, as pointed out above, a few professors that had been appointed to committees were officially replaced be another randomly chosen professor. Second, according to anecdotal evidence, some professors did not attend the exam (or part of it) without a proper justification. Unfortunately, we are unable to observe evaluators' attendance. In what follows we measure committees' composition using the outcome of the random draw. Therefore, our analysis provides the intention-to-treat effect.

4.1.1 Linear model

We estimate the following linear probability model:[15]

$$y_{ie} = \beta_0 + \beta_1 f_i + \beta_2(s_e - \mu_e) + \beta_3 f_i(s_e - \mu_e) + \beta_4 z_e + \epsilon_{ie} \quad (1)$$

where y_{ie} indicates whether individual i qualified in exam e, f_i is a dummy variable that takes value one if the candidate is female, s_e and μ_e represent respectively the actual and the expected proportion of female evaluators in the committee, and z_e is the number of available positions per candidate. We cluster standard errors by exam and by applicant to account for the fact that the performance of an individual in a given exam may depend on the performance of other candidates in the exam, or on her own performance in some other exam (see Cameron et al. 2006).

To facilitate the interpretation of coefficients we have rescaled z_e by subtracting its sample mean. Thus, coefficient β_0 reflects the average probability of promotion of male candidates and β_1 captures the difference in the success rate of female and male candidates. Coefficient β_2 indicates how the success rate of male candidates is affected by an increase in the proportion of female evaluators, and β_3 indicates how the gender gap in promotions is affected by variations in the gender composition of the committee.

We estimate equation (1) for exams to full professor positions and for exams to associate professor positions separately. Results for full professor exams are reported in the upper panel of Table 5. The average success rate for male applicants to full professor positions is 10.2%, about 1.3 percentage points higher than the success rate for female applicants. Since the specification does not include controls for candidate quality, this gap may reflect both the existence of gender differences in candidates' quality and discrimination by evaluators. The gender gap in promotion is lower in committees with a relatively higher proportion of female evaluators. Each additional female evaluator decreases the chances of success of male applicants by 6%, and increases the chances of success of female applicants by 14%.[16] The number of available positions per candidate

[15]Results 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.

[16]The average success of male candidates in full professor exams is 0.102. In a committee with

has a very strong effect on candidates' chances of being promoted, and the magnitude of the coefficient is very close to one. This is consistent with the fact that practically all positions are filled.

These results may potentially hide significant differences across disciplines, perhaps related to their degree of feminization (Graves and Powell 1995). In columns 2 to 7, we replicate the analysis for six broadly defined groups of disciplines, ordered according to their degree of feminization. We do not observe any clear association with the degree of feminization of disciplines. Only in Medicine do we find that the effect is significantly different from zero at standard levels, but it is not possible to reject the hypothesis that the effect is similar across all disciplines.

In the lower panel of Table 5 we report results from exams to associate professor positions. As in exams to full professor positions, we observe that men are more successful than women: 11.9% of male candidates are promoted, compared to only 9.9% of female candidates. In this case, an increase in the proportion of female evaluators negatively affects the chances of success of female applicants, though this effect is only significant at the 10% level. In quantitative terms, each additional female committee member increases the gender gap by 0.8 percentage points. In terms of the success rate of candidates, each additional female evaluator is associated with a 2% increase in the chances of promotion of male candidates, and a 5% decrease in the chances of promotion of female candidates. The effect of committees' gender composition is statistically similar across scientific areas and, again, it does not seem to be related to the degree of feminization of the disciplinary area. However, in none of the disciplinary areas can we reject the possibility that the effect is equal to zero.[17]

In sum, the gender composition of committees affects the outcome of promotion

seven members, each additional female evaluator decreases male candidates' chances of success by $(-0.042/7) * (1/0.102) * 100\% \approx -6\%$. The success rate of female candidates is equal to $0.102 - 0.013 = 0.089$, hence each additional female evaluator increases female candidates' chances of success by $[(-0.042 + 0.127)/7] * (1/0.089) * 100\% \approx 14\%$.

[17]We have also divided disciplines in two groups depending on whether the proportion of tenured professors in the discipline who are female is below or above the median (28% of tenured professors). The effect of committees' gender composition does not seem to vary with the degree of feminization of the field, neither in exams to associate professor positions nor in exams to full professor positions (results not reported). We obtain very similar results if we divide the sample according to the number of female professors in the field instead of the proportion.

decisions, but the direction and the magnitude of the effect depends on the seniority of the academic position. In exams to full professor positions, female candidates benefit strongly from a larger presence of women in the evaluation committee. In contrast, in exams to associate professor positions, the gender composition of committees is only marginally related to candidates' success. If anything, in this case, a larger proportion of female evaluators reduces the chances of promotion of female applicants.

4.1.2 The rank of evaluators

The effect of the gender composition of committees varies across different types of exams. This might potentially reflect the different composition of committees in terms of evaluators' rank. As explained in Section 2, in exams to full professor positions, committees include only full professors, whereas in exams to associate professor positions, three evaluators are full professors and four evaluators are associate professors. Thus, the results above could potentially reflect differences in the behavior of evaluators associated with their rank. In Table 6 we analyze how promotion decisions in exams to associate professor positions are affected by (i) the proportion of female full professors among evaluators, and (ii) the proportion of female associate professors among evaluators. Both have a negative effect on the success rate of female candidates, but none of these effects is statistically different from zero at standard levels. Most importantly, the effect of a larger presence of female full professors in exams to associate professor positions (Table 6) is significantly different from its effect in exams to full professor positions (column 1, upper panel of Table 5).[18] This result is consistent with evaluators having different gender attitudes depending on the position at stake.

4.1.3 Other committee characteristics

We have observed that the gender of evaluators significantly affects promotion decisions to full professor positions and it might also marginally affect promotions to associate professor positions. We now investigate whether these results reflect the fact

[18] A Wald test rejects the equality of these two coefficients at the 5% level.

that female and male evaluators are different in some dimension that could be affecting evaluations. To deal with this issue we estimate equation (1) including the age of evaluators, their tenure, and their research production. In order to exploit only exogenous variations in these variables, we use as independent variable the difference between the expected and the actual value of each committee characteristic. We do not find any significant effect of evaluators' age or tenure on the chances of success of female and male applicants (Table 7, columns 1 and 3). The inclusion of these controls does not significantly affect the estimated effect of the gender composition of committees.

Another characteristic that may be relevant is the research quality of evaluators. Descriptive evidence presented in Table 2 suggests that eligible female professors tend to have slightly lower research productivity as measured by the number of publications and citations per publication. Next we add controls for evaluators' research quality to our previous specification. In particular, we control for the normalized number of ISI publications (weighted by the number of co-authors) and the normalized number of received citations per publication. Given that research output is not defined for professors in the Humanities, we estimate this model excluding exams in this disciplinary area. Results for full and associate professor exams are shown in columns 2 and 4 of Table 7 respectively. The research quality of evaluators does not affect the gender mix of successful candidates. The introduction of these controls does not significantly affect our previous estimates.

4.1.4 Nonlinearities

Equation (1) assumes that the effect of gender on candidates' chances of promotion is linear. Is this assumption justified? Nonlinearities could arise for several reasons. First, the presence of a woman in the committee may affect the voting behavior of male evaluators. 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 majority basis. Therefore, having a committee where the 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 the different possible gender compositions arise as the result of the random draw. Using information on the gender mix in the pool of eligible evaluators, it is possible to calculate the probability that exactly j female evaluators were drawn in each exam, $p(d_{je})$.[19] The following model allows for the gender composition of the committee to have a nonlinear effect on candidates' success rate:

$$\begin{aligned} y_{ie} = \gamma_0 + \sum_{j=1}^{7} \gamma_j d_{je} + \lambda_0 f_i + \sum_{j=1}^{7} \lambda_j f_i d_{je} \\ + \sum_{j=1}^{7} \delta_j p(d_{je}) + \sum_{j=1}^{7} \mu_j f_i p(d_{je}) + \nu z_e + \epsilon_{ie} \end{aligned} \tag{2}$$

where d_{je} is a dummy that takes value one if the number of female evaluators in exam e is equal to j. Given the scarce number of committees with four, five or six female evaluators, we aggregate these committees in a single group.

We report the results from estimating equation (2) in columns 1 and 4 of Table 8. Both in full professor and associate professor exams, the linearity of the effect cannot be rejected by the data. If anything, increases from two to three female evaluators seem to have a slightly weaker effect, but the estimation is not accurate enough to make statistical claims.

4.2 Which committees discriminate?

The identification strategy above, which relies on exploiting a random lottery, allows us to estimate consistently how variations in the gender composition of evaluation committees affect the chances of success of male and female applicants. However, in that setup it is not possible to know which committees are biased. To answer this question, we need to control for candidates' quality. Below we analyze how the chances of success varies for candidates with similar (observable) quality across committees

[19] Details about how these probabilities are calculated are provided in Appendix C.

with different gender composition. Thus, we estimate the following model:

$$y_{ie} = \gamma'_0 + \sum_{j=1}^{7} \gamma'_j d_{je} + \lambda'_0 f_i + \sum_{j=1}^{7} \lambda'_j f_i d_{je}$$
$$+ \sum_{j=1}^{7} \delta'_j p(d_{je}) + \sum_{j=1}^{7} \mu'_j f_i p(d_{je}) + \nu' z_e + \eta q_{ie} + \epsilon_{ie} \quad (3)$$

where q_{ie} represents candidates' quality. We proxy for quality using ISI publications and received citations before applying to exam e. According to survey information, publications in journals covered by the ISI Web of Science are considered by Spanish professors to be the most important criterion for promotion decisions (Buela-Casal and Sierra 2006). Additionally, we control for candidates' age. Conditional on having produced the same research output, relatively younger candidates might be considered more productive than their older counterparts. Candidates' research production and age have been standardized to have zero mean and unit standard deviation within each exam. As long as q_{ie} captures all gender differences in candidates' quality, $\lambda'_j, j \in \{0, \ldots, 7\}$, it provides information on whether female candidates are being discriminated against by committees with j female evaluators.

Since the ISI information does not capture the research production in the Humanities accurately, we exclude exams in this disciplinary area from the analysis. This reduces the number of observations by approximately one third. For the sake of comparison, we reestimate model (2) for the sample excluding Humanities. We also exclude candidates for whom age information was not available. In columns 2 and 5 of Table 8 we consider respectively exams to full professor positions and exams to associate professor positions. The magnitude of the coefficients is similar to our previous results, but standard errors become larger.

In columns 3 and 6 we provide the results from estimating equation (3) on the same sample. As expected, we find that individuals who publish more, and whose publications receive more citations, are significantly more likely to be promoted. In addition, conditional on their research production, younger applicants to associate professor positions have higher chances of success. Controlling for candidates' observable quality does not significantly affect the estimated effect of the gender composition of committees. In Figure 2 we plot the corresponding point estimates. In committees where all evaluators are male, female applicants to full professor positions are significantly less likely to be promoted than (apparently) equally qualified male applicants. In committees with at least one female evaluator, female and male applicants have similar chances. In the case of exams to associate professor positions, women have the same chances of promotion as men in all committees apart from committees with a majority of female evaluators. In this case, female candidates have significantly lower chances of success than male candidates of similar research quality.

Given that the identification strategy relies on observables, the consistency of these results should be considered with caution. First, one may argue that, if promotion decisions are gender biased, a similar bias may be present in the publication process.[20] Second, women and men might differ in some relevant dimension other than publication record. This might be an important issue in exams to associate professor positions, where lecturing ability was also evaluated. This ability might be systematically different for female and male candidates, even after taking into account research output.

4.3 Which committee members discriminate?

We observe the final decisions taken by committees but, unfortunately, we cannot observe the evaluations that were cast by each individual committee member. Committee members discuss their evaluations before voting, then promotion decisions are made by the committee on a majority basis.

In exams to full professor positions, the evidence suggests that, based on the observable information on candidates' research production, all-male committees discriminate against female candidates. Committees which include at least one female evaluator seem to treat male and female candidates equally. This evidence is consistent with male evaluators discriminating against female candidates when there are no women in

[20]Blank (1991) conducted a randomized experiment at The American Economic Review and found that female-authored papers are relatively less likely to be published compared to male-authored papers when evaluators observe the authors' identity, even though this difference is not significant at standard levels.

the committee, but not doing so in the presence of female evaluators.

In exams to associate professor positions results differ. If anything, our data suggests that, in committees where the majority of evaluators are male, equally qualified male and female candidates have similar chances of being promoted. Female majority committees discriminate against female candidates (or, equivalently, favor male candidates). This evidence is consistent with at least two hypotheses. First, female evaluators discriminate against female candidates. Second, male evaluators favor male candidates when sitting in committees with a female majority. The latter may arise if male committee members' identities are strengthened with the presence of female members in the committee (Akerlof and Kranton 2000). Nevertheless, it seems reasonable to expect female evaluators to be able to have a larger influence on promotion decisions when they are majority among evaluators. Note that, as discussed above, given the absence of information on candidates' teaching quality, our analysis discussing which committees might be discriminating in associate professor exams should be considered with caution.

5 Interpretation of results

The evidence suggests that, in exams to full professor positions, male evaluators are relatively less favorable toward female candidates. However, when we move down the academic career ladder, the gender composition of the committees is not relevant or, if anything, the effect of gender composition of the committees is of the opposite sign.

Below, we discuss several theories that are consistent with our findings.

5.1 Differences in evaluation or differences in performance?

At least part of the observed effect could potentially reflect the existence of self-fulfilling expectations. Applicants can observe the composition of committees before taking the exam; hence, the gender composition of the committee might affect promotions by affecting the behavior of candidates.

For instance, some candidates may decide not to take the exam if, given the observed committee composition, their expected probability of being promoted is not high enough to compensate for the cost of attending the exam. Unfortunately, we cannot observe which applicants actually participated in the exam; therefore we cannot test directly whether committee composition affects participation. However, if the effect of committee composition is driven by participation decisions, this effect should be stronger for candidates for whom the cost of participation is higher. Candidates with a very low cost of attending the exam will probably take the exam regardless of the composition of the committees.

We proxy for the cost of attending the exam by the approximate car travel time between the city where the exam was held and the university where the candidate is based.[21] We compare the effect of committees' gender composition on candidates based within a four-and-a-half-hour driving distance from the exam (the median distance), and the effect of committees' gender composition on candidates residing further away. Given that we do not observe the affiliation of candidates to associate professor positions, we can only perform this analysis for candidates to full professor positions. As shown in Table 9, there are no differences in the effect of committee gender composition across the two groups of candidates. This result is consistent with the idea that the gender composition of committees does not significantly affect the participation decisions of candidates that have some real chance of being promoted.

The composition of the committee could also affect the performance of candidates during the exam. For instance, according to the "stereotype threat" hypothesis, in domains in which women are already negatively stereotyped, interacting with a sexist man can trigger a social identity threat, undermining women's performance (Steele 1997).

In theory, a potential "stereotype threat" is more likely to occur in exams to associate professor positions than in exams to full professor. In the former, candidates are relatively less experienced and their performance might be more strongly affected by

[21]As calculated by `http://www.ViaMichelin.com`, retrieved in January 2010.

the attitude of the committee. Moreover, in addition to the two stages in full professor exams, candidates to associate professor positions must give a lecture on one topic of the syllabus of an undergraduate course. However, the evidence shows that female applicants to associate professor positions are, if anything, relatively more successful, not less, when evaluated by committees with more men. This is at odds with the "stereotype threat" hypothesis. In full professor exams, we do not observe performance and thus we cannot completely discard the existence of a "stereotype threat". Nonetheless, given the experience of candidates and the format of the evaluation it does not seem very likely that this hypothesis is important.

In sum, the evidence is consistent with the idea that the gender composition of committees affects promotions directly through committee members' evaluations and it has, if any, a limited impact on candidates' performance. In particular, men are less favorable toward women when they decide on promotions to full professor positions.

5.2 Why are male evaluators relatively less favorable toward female candidates applying to full professor positions?

There are at least four explanations that are consistent with the behavior of male evaluators.

First, this behavior might reflect the existence of information asymmetries across genders. If, for some reason, these information asymmetries are relevant in exams to full professor positions (and not in exams to associate professor positions), that would explain why male evaluators are relatively less favorable toward female candidates in the former. Information asymmetries across genders can arise for a number of reasons. It has been shown that men and women tend to specialize in different research subfields (Almunia et al. 2005). If evaluating ability in a given dimension requires some related knowledge, evaluators will tend to prefer candidates of their own gender (Bagues and Perez-Villadoniga 2008). Full professor positions involve important decision making duties, that are absent in associate professor positions. Evaluators of different gender may have different priors relative to how men and women perform in these tasks.

It seems natural to think that information asymmetries should be less severe in smaller disciplinary areas where individuals are more likely to have interacted previously. Additionally, in larger areas, the potential number of subfields is larger, and thus the probability that evaluators are acquainted with candidates' research is lower. Moreover, the number of candidates tends to be larger, which implies that evaluators potentially have less time to evaluate each candidate. We define a large discipline as a discipline where the number of tenured professors is larger than the median number of tenured professors (two hundred seventy eight professors). We then estimate equation (1) for the subsamples of applicants in large and small disciplines. We observe that the gender composition of committees only matters in small disciplines. In large disciplines, the chances of success of female and male candidates do not depend on the gender composition of committees (columns 1 and 2 of Table 10). This suggests that information asymmetries are not the source of the observed gender bias.

Second, male evaluators may have some preference for relatively younger women because of some sort of "beauty and the labor market" story (Hamermesh and Biddle 1994). That would explain why male evaluators discriminate against female candidates in exams to full professor positions, but not in exams to associate professor positions. If this hypothesis holds, older female candidates should experience a relative disadvantage from being evaluated by committees with more men. We reestimate our model allowing the effect of committee composition to differ across female and male candidates belonging to different age groups. We do not find any clear pattern associated to candidates' age (Figure 3). In exams to full professor positions, an increase in the number of men in the committee has, if anything, a stronger negative effect on the success rate of relatively younger female candidates.

Third, discrimination against female candidates could arise as a consequence of cronyism: if evaluators tend to favor their friends, and friends tend to be of the same gender, this would result in candidates from the same gender as evaluators being favored.[22] If, for some reason, networks are more important at later stages on the

[22]Boschini and Sjögren (2007) study the co-authorship patterns of economists and find that there is a preference for having co-authors of the same gender.

academic career, this would explain why male evaluators are relatively more favorable toward male applicants in exams to full professor positions. We cannot directly observe individuals' social and professional networks in our sample. Nevertheless, in full professor exams we can observe the affiliation of applicants and evaluators. Affiliation captures probably one of the most important platforms for building networks.

In order to know whether this effect is gendered, we analyze how promotion decisions are affected by (i) the proportion of female evaluators with the same affiliation as the candidate's, (ii) the proportion of male evaluators with the same affiliation as the candidate's and, finally, (iii) the proportion of female evaluators in the committee coming from an institution different from the candidate's. As shown in Table 11, the effect of gender composition on candidates' chances of success is in fact driven by evaluators from other institutions. Male and female evaluators favor candidates' from their own institution, but this effect does not depend on gender. In quantitative terms, the presence of a colleague on the committee increases male and female applicants' chances of being promoted by approximately four percentage points.

Finally, the evidence may reflect the existence of ambivalent sexism. According to this theory, sexism might be a "multidimensional construct that encompasses two sets of sexist attitudes: hostile and benevolent sexism" (Glick and Fiske 1996). Male evaluators might experience sexist antipathy towards female candidates applying to top academic positions, but not to toward female candidates applying to lower-level positions. This would explain why male evaluators treat female candidates differently depending on the position at stake.

6 Conclusions

In the last few decades there has been a significant increase in the number of women starting academic careers. Currently women account for about half of PhD graduates, but the larger presence of women at the lower rungs of the academic ladder has not translated into proportional increases in the presence of women at the top.

In order to increase the representation of women in top positions in academia, some countries are mandating gender quotas in hiring and promotion committees. The motivation underlying the imposition of these gender quotas is the perception that the persistence of the glass ceiling may be due to (male) discrimination against women. According to this view, increasing the number of women who sit on evaluation committees would improve the career opportunities of women. This paper studies whether this policy may work.

We exploit evidence from a large-scale randomized natural experiment: the system of centralized examinations that was implemented between 2002 and 2006 in Spain to determine promotion to associate and full professor positions. These evaluations involved around 30,000 applications and 7,000 evaluators in all academic disciplines. The fact that committee members were selected through a lottery allows us to consistently estimate the effect of committees' gender composition on promotions.

We find that the gender composition of committees is an important determinant of promotion, but the effect depends on the position at stake. In exams to full professor positions, the gender composition of committees has a strong effect on female and male candidates' chances of success. Each additional male evaluator decreases the number of successful female applicants by 14%, and it increases the number of successful male applicants by 6%. However, when these committee members decide on promotions to associate professor positions, no significant interaction between evaluators' and candidates' gender could be observed. If anything, female candidates to associate professor positions have higher chances in the committees composed of relatively more men. The results do not depend on the degree of feminization of the discipline, and they are not due to omitted characteristics of evaluators such as age, tenure or research quality.

In order to explore the source of the gender bias in the full professor exams, we have collected information on candidates' publication records from the ISI Web of Science. As long as this proxy captures all relevant gender differences in quality, we can ascertain which committees discriminate. Relying on this assumption, we find that in all-male committees, female candidates have significantly lower chances of success than equally qualified male candidates. In committees with at least one woman on the committee, we do not find significant differences between female and male candidates.

We find this gender bias to be evident in small disciplines, but not in large disciplines. This suggests that information asymmetries across evaluators is not likely to be the source of the gender bias. In fact, information asymmetries should be stronger in larger disciplines, where evaluators are less likely to be acquainted with candidates before the exam. Our findings also rule out that female candidates are discriminated against on the basis of age, and similarly, we can rule out the existence of gendered networks. The observed gender bias may reflect some type of ambivalent sexism: male evaluators might oppose female candidates applying to top academic positions, while they do not object to female candidates applying to lower-level positions.

Our findings have direct policy implications for countries and institutions that encourage female representation on hiring and promotion committees. In the case of Spain, since 2007, the gender quotas exist in all hiring and promotion committees in public institutions, including academic committees. The quota mandates that at least 40% of committee members must be of each gender. Given the low numbers of women in the upper levels of academia, this has implied that women participate in committees much more often than men. A back-of-the-envelope calculation suggests that a female full professor will sit in committees four times more often than a male full professor.[23] Sitting on committees reduces the available time for research; thus gender quotas might lower the productivity of women who have managed to overcome the glass ceiling.[24] Since, as shown in this paper, quotas may have a positive effect on female promotion to full professor positions only, our work provides strong evidence against implementing gender quotas at the lower rungs of the academic ladder. Moreover, according to our findings, the presence of a single female on the committee is enough to overcome the gender bias in evaluation.

[23] This figure is the result of considering that the proportion of women in the population of full professors is equal to 15%, and that at least 40% of committee members should be women $[(0.85 - (0.85 * 0.60))/(0.15 * 0.60) \approx 4]$.

[24] Daniel Hamermesh warns young female economists to avoid requests to sit in committees 'like the plague'. According to his view, asking women disproportionately to sit on committees constitutes 'another form of sexual exploitation' (An Old Male Economist's Advice to Young Female Economists, CSWEP Newsletter, Winter 2005, p. 2).

References

[1] Abrevaya, Jason and Daniel S. Hamermesh (2011), "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?," *Review of Economic and Statistics*, forthcoming.

[2] Akerlof, George A. and Rachel E. Kranton (2000), "Economics And Identity," *The Quarterly Journal of Economics*, MIT Press, Vol. 115(3), pp. 715-753.

[3] Almunia, Miguel, Juan Jose Dolado and Florentino Felgueroso (2005), "Do Men and Women Economists Choose the Same Research Fields? Evidence From Top-50 Departments," *CEPR Discussion Paper* 5421.

[4] 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. 1301-28.

[5] Bagues, Manuel and Maria Jose Perez-Villadoniga (2008), "Why do I like people like me?," *Universidad Carlos III Working Paper* 2008-06.

[6] 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

[7] Blank, Rebecca M. (1991), "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review," *American Economic Review*, Vol. 81(5), pp. 1041-67.

[8] 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.

[9] 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.

[10] Broder, Ivy E. (1993), "Review of NSF Economics Proposals: Gender and Institutional Patterns," *American Economic Review*, Vol. 83(4), pp. 964-970.

[11] Buela-Casal, Gualberto and Juan Carlos Sierra (2006), "Criterios, indicadores y estándares para la acreditación de Profesores Titulares y Catedráticos de Universidad, Psicothema," Vol. 19(4), pp. 537-551.

[12] Cameron, Colin A., Jonah B. Gelbach and Douglas L. Miller (2006), "Robust Inference with Multi-way Clustering," *NBER Technical Working Papers* 0327.

[13] Combes, Pierre-Philippe, Laurent Linnemer and Michael Visser (2008), "Publish or peer-rich? The role of skills and networks in hiring economics professors," *Labour Economics*, Vol. 15(3), pp. 423-441.

[14] Cruz-Castro, Laura, Luis Sanz-Menéndez and Jaime Aja Valle (2006), "Las trayectorias profesionales y académicas de los profesores de universidad y los investigadores del CSIC," *Unidad de políticas comparadas (CSIC) Working Paper* 06-08.

[15] European Commission (2008), "Mapping the Maze: Getting More Women to the Top in Research," Luxembourg: Publication Office of the European Union.

[16] European Commission (2009), "She Figures 2009: Statistics and Indicators on Gender Equality in Science," Luxembourg: Publication Office of the European Union.

[17] Ginther, Donna K. and Kathy J. Hayes (2003), "Gender Differences in Salary and Promotion for Faculty in the Humanities, 1977-1995", *Journal of Human Resources*, Vol. 38(1), pp. 34-73.

[18] 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.

[19] 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.

[20] Glick, Peter and Susan T. Fiske (1996), "The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism," *Journal of Personality and Social Psychology*, Vol. 70(3), pp. 491-512.

[21] Graves, Laura. M. and Gary. N. Powell (1995), "The Effect of Sex Similarity on Recruiters' Evaluations of Actual Applicants: A Test of the Similarity-Attraction Paradigm," *Personnel Psychology*, Vol. 48(1), pp. 85-98.

[22] Hamermesh, Daniel and Jeff E. Biddle (1994), "Beauty and the Labor Market," *American Economic Review*, Vol. 84(5), pp. 1174-94.

[23] Holmes, Mary Anne and Suzanne O'Connell (2007), "Leaks in the pipeline," *Nature*, Vol. 446, pp. 346.

[24] Instituto Nacional de Estadística, Estadística de la Enseñanza Universitaria, several issues.

[25] McDowell, John M.; Larry D. Singell, Jr. and James P. Ziliak (1999), "Cracks in the Glass Ceiling: Gender and Promotion in the Economics Profession," *American Economic Review, Papers and Proceedings*, Vol. 89(2), pp. 392-396.

[26] National Research Council (2007), *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering*, Washington D.C.: The National Academy Press.

[27] National Research Council (2010), *Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty*, Washington D.C.: The National Academy Press.

[28] National Science Foundation (2009), "Characteristics of Doctoral Scientists and Engineers in the United States: 2006. Detailed Statistical Tables NSF 09-317," Arlington, VA. Available at `http://www.nsf.gov/statistics/nsf09317/`.

[29] Steele, Claude M. (1997), "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist*, Vol. 52(6), pp. 613-29.

[30] Tesch, Bonnie; Helen M. Wood; Amy L. Helwig; Ann Butler Nattinger (1995), "Promotion of Women Physicians in Academic Medicine: Glass Ceiling or Sticky Floor," *Journal of the American Medical Association*, 273(13), pp. 1022-1025.

Table 1: Descriptive statistics – Examinations

		1	2	3	4	5	6	7
		Full professor exams						
		All	Engineering	Math and Physics	Medicine	Social Sciences	Biology and Chemistry	Humanities
Number of exams		436	37	45	54	56	83	161
Positions per exam	Mean	3.00	3.00	3.35	2.59	3.16	3.35	2.80
	Std. Dev	1.85	2.09	2.62	1.50	1.81	2.10	1.44
	Min	1	1	1	1	1	1	1
	Max	12	11	12	10	10	10	8
Candidates per exam	Mean	30.33	32.81	41.71	27.28	28.63	30.25	28.24
	Std. Dev	18.85	17.34	23.71	16.47	16.85	19.65	17.65
	Min	6	15	13	8	12	6	6
	Max	134	87	134	88	118	113	119
Positions per candidate	Mean	0.12	0.09	0.08	0.11	0.12	0.13	0.12
	Std. Dev	0.06	0.04	0.04	0.06	0.05	0.07	0.07
	Min	0.02	0.04	0.02	0.04	0.04	0.03	0.02
	Max	0.38	0.20	0.18	0.30	0.28	0.33	0.38
Proportion of positions filled	Mean	0.98	1	1	1	0.99	0.97	0.98
	Std. Dev	0.09	0	0	0	0.04	0.12	0.11
	Min	0	1	1	1	0.75	0.50	0
	Max	1	1	1	1	1	1	1
		Associate professor exams						
		All	Engineering	Math and Physics	Medicine	Social Sciences	Biology and Chemistry	Humanities
Number of exams		455	73	47	64	53	46	172
Positions per exam	Mean	4.79	4.78	5.55	3.63	5.75	6.35	4.30
	Std. Dev	4.75	4.66	5.88	2.92	6.19	6.55	3.66
	Min	1	1	1	1	1	1	1
	Max	25	25	25	17	25	25	23
Candidates per exam	Mean	41.30	41.48	69.36	36.59	50.74	41.93	32.23
	Std. Dev	35.83	29.38	58.47	35.54	39.54	33.52	23.52
	Min	3	4	7	7	9	8	3
	Max	274	145	274	174	213	168	146
Positions per candidate	Mean	0.13	0.12	0.09	0.13	0.12	0.15	0.15
	Std. Dev	0.08	0.08	0.07	0.06	0.06	0.07	0.09
	Min	0.02	0.03	0.02	0.02	0.03	0.04	0.03
	Max	0.67	0.50	0.31	0.29	0.33	0.27	0.67
Proportion of positions filled	Mean	0.96	0.97	0.99	0.96	0.94	0.93	0.96
	Std. Dev	0.15	0.13	0.03	0.16	0.16	0.14	0.16
	Min	0	0.33	0.83	0	0.3	0.50	0
	Max	1	1	1	1	1	1	1
Number of exams		455	73	47	64	53	46	172

Table 2: Descriptive statistics – Eligible evaluators and candidacies

	1	2	3	4	5	6
	Eligible Evaluators					
	Full professors			**Associate professors**		
	Total	Male	Female	Total	Male	Female
	Total	Male	Female	Total	Male	Female
Age	52.88	52.93	52.57	44.95	44.87	45.09
	(0.09)	(0.09)	(0.20)	(0.06)	(0.08)	(0.10)
	[42885]	[36614]	[6271]	[58770]	[38170]	[20600]
Tenure in position	12.92	13.41	9.97	10.34	10.27	10.48
	(0.10)	(0.11)	(0.21)	(0.05)	(0.06)	(0.08)
	[47585]	[40787]	[6798]	[60278]	[39226]	[21052]
Publications, weighted by co-authors	9.98	10.22	8.30	5.86	6.03	5.50
	(0.17)	(0.19)	(0.36)	(0.05)	(0.07)	(0.08)
	[33744]	[29463]	[4281]	[43217]	[29377]	[13840]
Citations per publication	11.47	11.43	11.77	10.28	9.96	10.96
	(0.14)	(0.15)	(0.33)	(0.09)	(0.11)	(0.14)
	[33744]	[29463]	[4281]	[43217]	[29377]	[13840]
Publications, weighted by co-authors, normalized	0.00	0.02	-0.16	0.00	0.04	-0.09
	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)
	[33744]	[29463]	[4281]	[43217]	[29377]	[13840]
Citations per publication, normalized	0.00	0.01	-0.05	0.00	0.02	-0.04
	(0.01)	(0.02)	(0.04)	(0.01)	(0.01)	(0.02)
	[33744]	[29463]	[4281]	[43217]	[29377]	[13840]
Total number of observations	47688	40859	6829	60729	39530	21199
Proportion of the sample		0.86	0.14		0.65	0.35
Total number of individuals	7909	6741	1168	21945	13897	8048
	Candidacies					
	Full professor exams			**Associate professor exams**		
	Total	Male	Female	Total	Male	Female
Age	46.39	46.29	46.64	37.35	37.68	36.85
	(0.10)	(0.12)	(0.18)	(0.08)	(0.11)	(0.12)
	[12350]	[8808]	[3542]	[17653]	[10635]	[7018]
Publications, weighted by co-authors	7.40	7.63	6.72	4.59	4.83	4.12
	(0.14)	(0.16)	(0.26)	(0.08)	(0.09)	(0.13)
	[8678]	[6487]	[2191]	[13248]	[8712]	[4536]
Citations per publication	11.20	11.11	11.47	9.00	8.98	9.05
	(0.18)	(0.21)	(0.31)	(0.15)	(0.21)	(0.18)
	[8678]	[6487]	[2191]	[13248]	[8712]	[4536]
Publications, weighted by co-authors, normalized	0.00	0.03	-0.10	0.00	0.02	-0.03
	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)
	[8678]	[6487]	[2191]	[13248]	[8712]	[4536]
Citations per publication, normalized	0.00	0.01	-0.03	0.00	0.00	0.00
	(0.02)	(0.02)	(0.04)	(0.01)	(0.02)	(0.02)
	[8678]	[6487]	[2191]	[13248]	[8712]	[4536]
Total number of observations	13224	9480	3744	18792	11351	7441
Proportion of the sample		0.72	0.28		0.60	0.40
Total number of individuals	6037	4255	1782	9952	5793	4159

Notes: Mean values, standard error in parentheses, number of observations for which statistics is computed in square brackets. Research data is not available in the Humanities. Normalized research indicators have zero mean and unit standard deviation among evaluators (candidates) in the same exam.

Table 3: Proportion of women among candidates and evaluators, by disciplinary area

	1	2	3	4	5	6	7
	All	Engineering	Math and Physics	Medicine	Social Sciences	Biology and Chemistry	Humanities
Full professors	0.14	0.05	0.08	0.12	0.16	0.17	0.18
Candidates to full professor positions	0.28	0.17	0.18	0.28	0.29	0.30	0.34
Associate professors	0.35	0.20	0.24	0.34	0.37	0.44	0.42
Candidates to associate professor positions	0.40	0.22	0.29	0.33	0.46	0.48	0.52

Table 4: Expected and actual committee composition

	1	2	3	4
	Full professor exams		Associate professor exams	
	Expected Committees	Actual Committees	Expected Committees	Actual Committees
Proportion of female evaluators	0.15	0.15	0.26	0.26
	(0.01)	(0.01)	(0.01)	(0.01)
Zero female evaluators	0.36	0.36	0.18	0.17
	(0.01)	(0.02)	(0.01)	(0.02)
One female evaluator	0.36	0.35	0.28	0.28
	(0.01)	(0.02)	(0.01)	(0.02)
Two female evaluators	0.19	0.18	0.25	0.27
	(0.01)	(0.02)	(0.01)	(0.02)
Three female evaluators	0.07	0.08	0.17	0.18
	(0.01)	(0.01)	(0.01)	(0.02)
Four or more female evaluators	0.03	0.03	0.12	0.10
	(0.00)	(0.01)	(0.01)	(0.01)
Age	53.12	53.17	48.73	48.87
	(0.10)	(0.15)	(0.14)	(0.18)
Tenure in position	13.28	13.25	11.65	11.69
	(0.10)	(0.17)	(0.10)	(0.15)
Publications, weighted by co-authors	8.91	8.40	6.64	6.58
	(0.36)	(0.38)	(0.23)	(0.27)
Citations per publication	12.27	11.66	11.13	11.12
	(0.28)	(0.32)	(0.25)	(0.32)

Notes: The sample includes 436 exams to full professor positions and 455 exams to associate professor positions. Standard errors are reported in parentheses. The calculation of the expected committee composition is explained in Appendix B.

Table 5: The effect of committees' gender composition on candidates' success

	1	2	3	4	5	6	7
	Full professor exams						
	All	Engineering	Math and Physics	Medicine	Social Sciences	Biology and Chemistry	Humanities
Female candidate	-0.013**	0.005	-0.012	-0.022	-0.020	-0.014	-0.013
	(0.005)	(0.020)	(0.013)	(0.019)	(0.016)	(0.014)	(0.008)
Proportion of female evaluators	-0.042***	-0.076	0.002	-0.119**	-0.061	0.002	-0.050*
	(0.016)	(0.063)	(0.030)	(0.047)	(0.048)	(0.040)	(0.026)
Female candidate*Proportion of female evaluators	0.127***	0.449	-0.009	0.347***	0.187	0.017	0.123*
	(0.046)	(0.342)	(0.139)	(0.127)	(0.123)	(0.111)	(0.071)
Positions per candidate	0.965***	1.017***	0.999***	1.001***	0.984***	0.956***	0.950***
	(0.003)	(0.085)	(0.002)	(0.003)	(0.014)	(0.011)	(0.016)
Constant	0.102***	0.091***	0.083***	0.101***	0.115***	0.113***	0.102***
	(0.001)	(0.004)	(0.002)	(0.005)	(0.005)	(0.004)	(0.003)
Adjusted R-squared	0.028	0.014	0.020	0.033	0.024	0.028	0.035
Number of observations	13224	1214	1877	1473	1603	2511	4546
	Associate professor exams						
	All	Engineering	Math and Physics	Medicine	Social Sciences	Biology and Chemistry	Humanities
Female candidate	-0.020***	-0.003	-0.009	-0.008	-0.025*	-0.058***	-0.023***
	(0.004)	(0.011)	(0.009)	(0.012)	(0.015)	(0.012)	(0.008)
Proportion of female evaluators	0.015	0.004	0.011	0.036	0.041	-0.013	0.035
	(0.012)	(0.024)	(0.009)	(0.049)	(0.042)	(0.044)	(0.028)
Female candidate*Proportion of female evaluators	-0.050*	-0.075	-0.053	-0.048	-0.090	-0.009	-0.072
	(0.026)	(0.096)	(0.036)	(0.108)	(0.082)	(0.089)	(0.048)
Positions per candidate	0.926***	0.943***	0.994***	1.014***	0.828***	0.879***	0.927***
	(0.022)	(0.026)	(0.018)	(0.042)	(0.143)	(0.049)	(0.022)
Constant	0.119***	0.113***	0.082***	0.098***	0.118***	0.168***	0.141***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.008)	(0.007)	(0.004)
Adjusted R-squared	0.033	0.030	0.030	0.034	0.024	0.035	0.030
Number of observations	18792	3028	3260	2342	2689	1929	5544

Notes: OLS estimates. Standard errors are clustered by applicant and by exam. 'Proportion of female evaluators' is the difference between the actual and the expected proportion of female evaluators in the committee. The number of positions per candidate is included among regressors and rescaled to have zero mean in the corresponding subsample.

Table 6: The effect of committees' gender composition on candidates' success, by evaluators' rank

	Associate professor exams
Female candidate	-0.020***
	(0.004)
Proportion of female full-professor evaluators	0.003
	(0.026)
Female candidate*Proportion of female full-professor evaluators	-0.023
	(0.058)
Proportion of female associate-professor evaluators	0.020
	(0.015)
Female candidate*Proportion of female associate-professor evaluators	-0.061*
	(0.033)
Constant	0.120***
	(0.002)
Adjusted R-squared	0.033
Number of observations	18792

Notes: OLS estimates. Standard errors are clustered by exam and by individual. 'Proportion of female evaluators of a given rank' is the difference between the actual and the expected proportion of female evaluators of this rank among all committee members. The number of positions per candidate is included in the regression and rescaled to have zero mean in the corresponding subsample.

Table 7: The effect of committees' composition on candidates' success, other characteristics

	1	2	3	4
	Full professor exams		Associate professor exams	
	All	Excluding Humanities	All	Excluding Humanities
Female candidate	-0.014***	-0.011	-0.020***	-0.019***
	(0.005)	(0.007)	(0.004)	(0.005)
Proportion of female evaluators	-0.039**	-0.037*	0.019*	0.014
	(0.016)	(0.021)	(0.011)	(0.012)
Female candidate*Proportion of female evaluators	0.116**	0.129**	-0.061**	-0.050
	(0.047)	(0.062)	(0.025)	(0.031)
Age of evaluators	-0.000	0.000	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Female candidate*Age of evaluators	0.001	-0.001	-0.000	-0.001
	(0.003)	(0.004)	(0.002)	(0.003)
Tenure of evaluators	0.001	0.000	0.001*	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)
Female candidate*Tenure of evaluators	-0.004	-0.002	-0.004*	-0.005*
	(0.002)	(0.003)	(0.002)	(0.002)
Publications of evaluators		-0.007		-0.008
		(0.006)		(0.005)
Female candidate*Publications of evaluators		0.027		0.020
		(0.025)		(0.014)
Citations per publication received by evaluators		-0.003		-0.008
		(0.006)		(0.005)
Female candidate*Citations per publication received by evaluators		0.015		0.004
		(0.020)		(0.015)
Constant	0.102***	0.101***	0.119***	0.111***
	(0.001)	(0.002)	(0.002)	(0.002)
Adjusted R-squared	0.028	0.025	0.033	0.033
Number of observations	13224	8678	18792	13248

Notes: OLS estimates. Standard errors are clustered by applicant and by exam. 'Proportion of female evaluators', 'Age of evaluators', 'Tenure of evaluators','Publications of evaluators' and 'Citations per publication received by evaluators' stand for the corresponding difference between the actual and the expected value for each committee characteristic. Missing values of evaluators' age and tenure were assumed to be equal to the average in the corresponding list of eligible evaluators of the same rank. Age and tenure are measured in years. Publications and citations have been standardized to have zero mean and one standard deviation within the corresponding pool of eligible evaluators. All regressions include the number of positions per candidate, which is rescaled to have zero mean in the corresponding subsample.

Table 8: The effect of committees' gender composition on candidates' success - Nonlinearities and Quality proxies

	1	2	3	4	5	6
	Full professor exams			Associate professor exams		
	All	Excluding Humanities		All	Excluding Humanities	
Female candidate	-0.032***	-0.033***	-0.028**	-0.002	0.001	-0.002
	(0.009)	(0.011)	(0.011)	(0.008)	(0.010)	(0.010)
One female evaluator	-0.005	-0.006	-0.006	0.006	0.009	0.008
	(0.003)	(0.005)	(0.005)	(0.004)	(0.006)	(0.006)
Two female evaluators	-0.011*	-0.007	-0.006	0.008*	0.006	0.005
	(0.006)	(0.010)	(0.010)	(0.005)	(0.005)	(0.005)
Three female evaluators	-0.009	-0.005	-0.003	0.001	0.003	0.003
	(0.010)	(0.013)	(0.013)	(0.008)	(0.012)	(0.009)
Four or more female evaluators	-0.037*	-0.044	-0.039	0.016*	0.019	0.019*
	(0.020)	(0.042)	(0.043)	(0.009)	(0.012)	(0.010)
Female candidate*One female evaluator	0.021*	0.031*	0.031*	-0.017	-0.025	-0.023
	(0.013)	(0.017)	(0.017)	(0.013)	(0.016)	(0.016)
Female candidate*Two female evaluators	0.037*	0.036	0.033	-0.024*	-0.020	-0.019
	(0.020)	(0.031)	(0.032)	(0.013)	(0.016)	(0.016)
Female candidate*Three female evaluators	0.033	0.019	0.014	-0.016	-0.029	-0.027
	(0.031)	(0.043)	(0.043)	(0.014)	(0.023)	(0.018)
Female candidate*Four or more female evaluators	0.088**	0.119*	0.111*	-0.047**	-0.046	-0.046**
	(0.044)	(0.066)	(0.065)	(0.018)	(0.030)	(0.021)
Publications			0.026***			0.017***
			(0.004)			(0.003)
Citations per publication			0.013***			0.012***
			(0.003)			(0.003)
Age			-0.005			-0.019***
			(0.003)			(0.003)
Constant	0.107***	0.108***	0.106***	0.114***	0.105***	0.105***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Adjusted R-squared	0.029	0.029	0.040	0.033	0.034	0.044
Number of observations	13224	8084	8084	18792	12433	12433

Notes: OLS estimates. Standard errors are clustered by exam and by individual. Probabilities to draw one, two, three, four, five, six or seven females, interactions of these probabilities with the female dummy, and the number of positions per candidate are included among regressors. All controls are rescaled to have zero mean in the corresponding subsample.

Table 9: The effect of committees' gender composition on candidates' success, by travel time

	1	2
	Full professor exams	
Travel time:	≤4.5 hours	>4.5 hours
Female candidate	-0.014*	-0.013*
	(0.008)	(0.008)
Proportion of female evaluators	-0.034	-0.049*
	(0.030)	(0.029)
Female candidate*Proportion of female evaluators	0.112*	0.137**
	(0.066)	(0.068)
Constant	0.106***	0.097***
	(0.003)	(0.003)
Adjusted R-squared	0.026	0.030
Number of observations	6406	6818

Notes: OLS estimates. Standard errors are clustered by exam and by individual. 'Proportion of female evaluators' is the difference between the actual and the expected proportion of female evaluators in the committee. Individuals traveling from abroad and from or to locations on the islands are included in column 2. The number of positions per candidate is included among regressors and rescaled to have zero mean in the corresponding subsample.

Table 10: The effect of committees' gender composition on candidates' success, by size of the discipline

	1	2	3	4
	Full-professor Exams		**Associate-Professor Exams**	
Discipline size:	≤median	>median	≤median	>median
Female candidate	-0.019**	-0.009	-0.019***	-0.022***
	(0.008)	(0.009)	(0.006)	(0.007)
Proportion of female evaluators	-0.067***	-0.020	0.017	0.021
	(0.025)	(0.021)	(0.021)	(0.014)
Female candidate*Proportion of female evaluators	0.195***	0.058	-0.055	-0.053
	(0.068)	(0.062)	(0.044)	(0.033)
Constant	0.116***	0.087***	0.138***	0.103***
	(0.002)	(0.002)	(0.003)	(0.003)
Adjusted R-squared	0.032	0.021	0.027	0.034
Number of observations	6757	6467	9331	9461

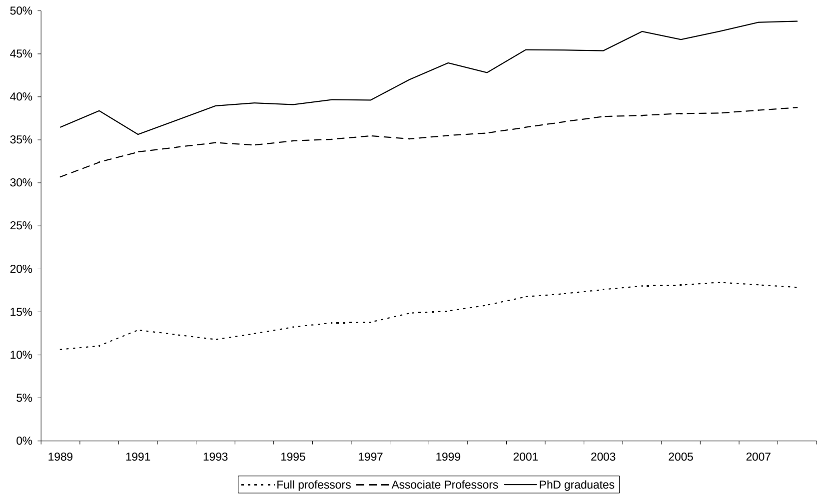
Notes: OLS estimates. Standard errors are clustered by exam and individual. 'Proportion of female evaluators' is the difference between the actual and the expected proportion of female evaluators in the committee. In the median discipline there are 278 tenured professors. The number of positions per candidate is included and rescaled to have zero mean in the corresponding subsample.

Table 11: The effect of committees' gender composition on candidates' success, by evaluators' affiliation

	Full professor exams
Female candidate	-0.014**
	(0.006)
Proportion of female evaluators from other universities	-0.043**
	(0.017)
Female candidate*Proportion of female evaluators from other universities	0.124***
	(0.046)
Proportion of female evaluators from candidate's university	0.285**
	(0.135)
Female candidate*Proportion of female evaluators from candidate's university	0.066
	(0.211)
Proportion of male evaluators from candidate's university	0.280***
	(0.044)
Female candidate*Proportion of male evaluators from candidate's university	-0.027
	(0.079)
Constant	0.102***
	(0.002)
Adjusted R-squared	0.035
Number of observations	13224

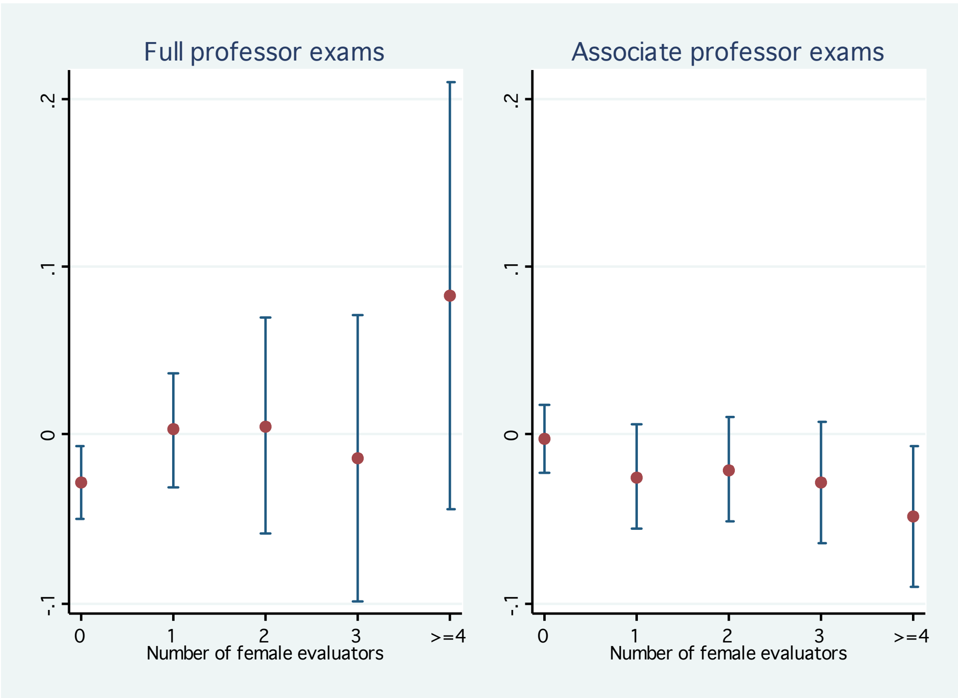
Notes: OLS estimates. Standard errors are clustered by exam and by individual. 'Proportion of female evaluators with a given affiliation' is the difference between the actual and the expected proportion of female evaluators with this affiliation among all committee members. The number of positions per candidate is included and rescaled to have zero mean in the corresponding subsample.

Figure 1: Proportion of women in Spanish Academia



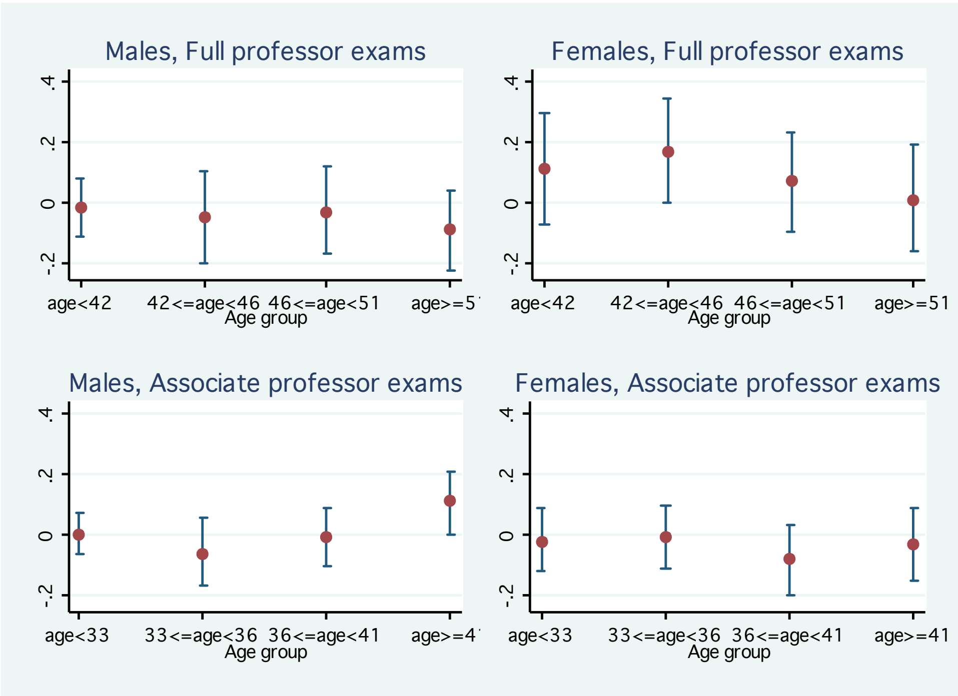
Source: Authors calculations based on information from Instituto Nacional de Estadística, Estadística de la Enseñanza Universitaria, several issues.

Figure 2: The effect of committees' gender composition on the gender promotion gap



Notes: The y-axis represents the difference in the success rate between female and male candidates, as estimated in Table 8. 95% confidence intervals are displayed.

Figure 3: The effect of committees' gender composition on the gender promotion gap, by candidates' age



Notes: The y-axis represents the difference in the success rate between female and male candidates. 95% confidence intervals are displayed.

Appendix A: Data Appendix

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 have also collected information on the research output of eligible evaluators and candidates from the ISI Web of Science.[25] Below we describe how the measures for gender, age, tenure, affiliation and research output were constructed.

Gender 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 on line for any personal picture or document that could allow to assign gender.

Age 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 numbers that they then assign 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 secretaries, which contains both the ID number and the date of birth. In 95% of theceses the assigned age is within a three year-interval of the actual age. In order to minimize potential errors, whenever

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

[26] There are a number of exceptions. For instance, this methodology will fail to identify the age of those individuals who obtained their nationality when they were older than 14. Still, immigration was a very rare phenomenon in Spain until the late 1990s. Additionally, some parents may have their kids 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.

our age proxy indicated that an (candidate to) associate professor is less than 27 years old and a (candidate to) full professor is less than 35 years old, we assign age a missing value (around 5% of the sample). In order to calculate the expected age and the actual average age of committee members, we assumed that the eligible evaluators, for whom the age proxy is missing, have the same age as other professors of the same academic rank in the same discipline. This proxy is not defined for non-Spaniards (less than 1% of the sample).

Affiliation and tenure in the position The Ministry provides information on affiliation and on tenure in the position for eligible evaluators. Given that most candidates to full professor positions are themselves eligible evaluators 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.

Research Output Information on scientific publications comes from Thompson ISI Web of Science (WoS). 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. WoS database includes over 10,000 high-impact journals in 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. Information was not collected for Humanities, as journal publications is probably a poor measure of quality in this field.

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 there are some surnames that are very common (i.e. Garcia, Fernandez, Gonzalez), and this may create homonymity problems. 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 with their maternal surname, or they may hyphenate the two surnames. As well, authors may 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.[27] 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 belongs to the same disciplinary category as the professor does, and the author's surname and initial, as reported by ISI, coincides (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) to (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 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 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 received by each publication subtracting the average number of citations received by Spanish-

[27]In 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.

authored articles published in the corresponding ISI disciplinary 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.

Appendix B: The Expected Share of Female Evaluators

In exams in which the list of eligible evaluators contains no more than one CSIC researcher and/or no more than one emeritus professor, the expected proportion of women in the committee is equal to the proportion of women in the list of eligible evaluators. However, as mentioned in Section 2, according to the design of the lottery no more than one CSIC researcher and no more than one emeritus professor was allowed to sit on a committee. In case that a second individual belonging to one of these categories was drawn, the draw was not considered. Therefore, if the list of eligible evaluators contains more than one CSIC researcher and/or more than one emeritus professor, the calculation of the expected proportion of females in the committee has to take into account the existence of the above rules. The probability that at least one researcher is drawn from the pool, p_R, and the probability that at least one emeritus professor is drawn, p_E, are equal to:

$$p_R = 1 - \frac{\binom{R}{0}\binom{P+E}{7-0}}{\binom{P+E+R}{7}}, \quad p_E = 1 - \frac{\binom{E}{0}\binom{P+R}{7-0}}{\binom{P+E+R}{7}}$$

where R is the number of researchers in the pool, E is the number of emeritus professors and P is the number of eligible professors that are not emeritus. Once these probabilities are computed, it is possible to calculate the expected proportion of female evaluators in exams to full professor positions:

$$\mu = \frac{1}{7}[p_R p_E(s_R + s_E + 5s_P) + p_E(1 - p_R)(s_E + 6s_P) \\ + p_R(1 - p_E)(s_R + 6s_P) + (1 - p_R)(1 - p_E)7s_P],$$

where s_j indicates the proportion of women in group j and $j \in \{R, E, P\}$.

Next, we calculate the expected proportion of female evaluators in exams to associate professor. The probability that at least one (junior) CSIC researcher is drawn from the pool of eligible evaluators in the category of associate professors, p_I, and the probability that at least one emeritus associate professor is drawn, p_O, is equal to:

$$p_I = 1 - \frac{\binom{I}{0}\binom{A+O}{7-0}}{\binom{A+O+I}{7}}, \quad p_O = 1 - \frac{\binom{E}{0}\binom{A+I}{7-0}}{\binom{A+O+I}{7}}$$

where I is the number of CSIC researchers in the pool, O is the number of emeritus associate professors and A is the number of associate non-emeritus professors. Then we can compute the expected proportion of women among those evaluators who are either full professors, or researchers or emeritus professors (drawn first)

$$\mu_1 = \frac{1}{3}[p_R p_E(s_R + s_E + s_P) + p_E(1 - p_R)(s_E + 2s_P)$$
$$+ p_R(1 - p_E)(s_R + 2s_P) + (1 - p_R)(1 - p_E)3s_P],$$

and the expected proportion of women among those evaluators who are either associate professors, or junior researchers or emeritus associate professors (drawn second)

$$\mu_2 = \frac{1}{4}[(1 - p_R)(1 - p_E) * [p_I p_O(s_I + s_O + 2s_A) + p_O(1 - p_I)(s_O + 3s_A)$$
$$+ p_I(1 - p_O)(s_I + 3s_A) + (1 - p_I)(1 - p_O)4s_A]$$
$$+ p_R(1 - p_E) * [p_O(s_O + 3s_A) + (1 - p_O)4s_A]$$
$$+ p_E(1 - p_R) * [p_I(s_I + 3s_A) + (1 - p_I)4s_A] + p_E p_R * 4s_A],$$

where s_j is the proportion of females among $j \in \{I, O, A\}$ type of eligible evaluators. Finally, we can compute the expected proportion of female evaluators in exams to associate professor:

$$\mu = \frac{1}{7}[3\mu_1 + 4\mu_2]$$

Following this methodology it is also possible to compute the expected committee composition in terms of any other observed individual characteristic of evaluators.

Appendix C: Nonlinearities

The probability to draw a certain number of women from the pool of potential evaluators in full professor exams is equal to:[28]

$$p(d) = \frac{\binom{F}{d}\binom{M}{7-d}}{\binom{F+M}{7}}, \quad \forall d \in \{0 \dots 7\}$$

where F is the total number of females among eligible evaluators in the corresponding exam list, and M is the number of males in the corresponding list.

For associate professor exams the probability of having zero, one, ..., or seven females in the committee is a bit more difficult to calculate. First, we compute the probability to draw zero, one, ..., or three females among full professor or senior researcher evaluators and zero, one, ..., or four females among associate professor or researcher evaluators:

$$p_1(d_1) = \frac{\binom{F_1}{d_1}\binom{M_1}{3-d_1}}{\binom{F_1+M_1}{3}}, \quad \forall d_1 \in \{0 \dots 3\}; \qquad p_2(d_2) = \frac{\binom{F_2}{d_2}\binom{M_2}{4-d_2}}{\binom{F_2+M_2}{4}}, \quad \forall d_2 \in \{0 \dots 4\},$$

where F_1 and F_2 are the numbers of females among eligible full professors and associate professors, respectively, and M_1 and M_2 are the numbers of males among eligible full professors and associate professors, respectively. Once these probabilities are computed, we calculate the probability to have zero, one, ..., or seven female evaluators in a given exam:

$$p(d) = \sum_{\substack{d_1, d_2: \\ d_1+d_2=d}} p_1(d_1) * p_2(d_2), \quad \forall d \in \{0 \dots 7\}$$

[28]For exams, in which the lists include more than one researcher and/or more than one emeritus professor, the weighting procedure described in Appendix B is applied to calculations presented below.