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

Natalia Zinovyeva [†]   Manuel Bagues [‡]

June 1st, 2011

Abstract

This paper studies how the gender composition of committees affects promotions. We exploit evidence from a large-scale randomized natural experiment: the system of centralized examinations to determine academic promotions that was implemented in Spain between 2002 and 2006. These competitions involved around 30,000 candidacies and 7,000 (randomly selected) evaluators. In exams to full professor positions, we find that evaluators tend to favor same-sex candidates who belong to their own academic network. This evidence is consistent with the existence of *old boys* and *old girls* networks. In exams to associate professor positions, senior evaluators do not exhibit any gender preference. Female junior evaluators exhibit an opposite-sex preference when assessing candidates from their own institution, perhaps for strategic reasons. Our results suggest that gender quotas may not necessarily increase female representation. Their effect will depend on the extent to which academic networks are gendered, evaluators' strategic concerns, and the position at stake.

Keywords: academic promotion, gender discrimination, randomized natural experiment.

JEL Classification: J71, J45.

[†]Corresponding author, Institute of Public Goods and Policies, Spanish National Research Council (IPP-CSIC), Madrid, email: natalia.zinovyeva@cchs.csic.es

[‡]Universidad Carlos III, Madrid, email: mfbagues@emp.uc3m.es

1 Introduction

Women have historically been under-represented in top academic positions.[1] In the past, this under-representation seemed to be due to the small number of women obtaining doctorates. In recent decades the number of women undertaking PhD studies has increased steadily, but the incidence of women in the upper echelons of the academic career ladder remains low, particularly among full professors. In Spain, in 2009 women represented 49% of PhD graduates, 39% of associate professors, but only 18% of full professors (Figure 1). The picture is qualitatively similar in the U.S. and Europe.[2]

There are several potential explanations for the persistent lack of women in top academic positions. Gender roles at the household level might make it more difficult for women to advance up the academic job 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 of academia (Holmes and O'Connell 2007). The lack of appropriate academic networks has also been identified as one of the potential causes of women's failure to advance in their career (Blau et al. 2010).

In addition to the above theories, it has been argued that the lack of women in top academic positions may be due to the absence of women in the evaluation committees that decide on promotions.[3] Beyond gender stereotypes and gender prejudices, male evaluators may potentially have a preference for same-sex candidates for a number of reasons. Men and women tend to specialize in different academic fields (Dolado et al. 2012, Hale and Regev 2011), if researchers tend to overrate the importance of their

---

[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, 2009) analyze the low presence of women in the high rungs of the academic career ladder 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).

[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.). Unfortunately, the report does not provide any empirical evidence to back this statement.

field of research, gender segregation might cause a gender bias in evaluations (Bagues and Perez-Villadoniga 2008). Evaluators also tend to overrate candidates in the same network (Zinovyeva 2011). If academic networks are gendered, female candidates might be at a disadvantage when evaluated by a male-majority committee.

However, there is no clear empirical evidence supporting the hypothesis that female evaluators are more favorable towards female candidates. The few papers that deal with the endogeneity of committee composition provide mixed results.[4] 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 of the interaction between the gender of referees and the gender of authors. In a non academic setting, Bagues and Esteve-Volart (2010) analyze hirings for entry-level positions in the Spanish Judiciary. The authors 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. In this case, the interaction between the gender of the applicant and the gender of the contact person in the hiring firm does not play any role. In sum, the evidence suggests that policy makers should not take female evaluators' alignment with female candidates as granted.

Given the persistence of a *glass ceiling*, we need a better understanding of the role of evaluators' gender. In this paper we address this issue using evidence from promotions in the Spanish public university system between 2002 and 2006.[5] During this period, all academic promotions were decided through nation-wide competitions. In total, ap-

---

[4]A neat empirical analysis is usually hard to come by because, in most situations, it is not possible to rule out the possibility that the gender composition of committees is related to the relative quality of female and male candidates. Another problem has been the lack of female evaluators. For instance, Blank (1991), Wennerås and Wold (1997) and Combes et al. (2008) study gender discrimination but cannot address whether the gender of evaluators matters due to the paucity of women among evaluators.

[5]In Spain, approximately 88% of university professors are based in public universities (Instituto Nacional de Estadística 2010).

proximately thirty thousand candidacies were evaluated by seven thousand (randomly chosen) evaluators. This setup has four exceptional features. First, evaluations were performed for two different positions along the career ladder: full professor positions and associate professor. Thus, it is possible to examine the relevance of gender in promotions at different levels, including top positions. Second, once the list of candidates was settled, evaluators were selected out of a pool of eligible professors using a lottery. Therefore, in these competitions, the gender composition of committees is independent of candidates' characteristics (conditional on the gender composition of the pool of eligible evaluators). This allows us to consistently estimate the causal effect of committees' gender composition on promotion decisions. Third, while most studies deal only with a single discipline, we examine promotions across all academic disciplines. Finally, we observe very detailed information on evaluators' and candidates' research productivity, their field of specialization and their academic networks. This information allows us to further investigate the origin of gender biases.

Our results are as follows. In exams to full professor positions, committees are composed of seven full professors. On average, an additional female evaluator increases the number of female applicants that are promoted by 14%. This result is driven by evaluators having a preference for same-sex candidates with whom they share research interests, particularly if they are in the same academic network (former student, co-author, same advisor,...). For instance, replacing a male evaluator with a female evaluator from the same academic network increases the chances of success of female candidates from this network by 78%. Evaluators' gender has no effect whatsoever on the chances of success of candidates with different research interests. This evidence is consistent with the existence of *old boys'* (and girls') networks.

In associate professor exams, committees include three full professors and four associate professors. On the one hand, and in contrast to exams to full professor positions, we do not observe any significant interaction between the gender of full professor evaluators and the success rate of candidates. This finding might potentially reflect that academic links at this stage are neither very strong nor very gendered. It might also be

due to the existence of ambivalent sexism, which arises when attitudes toward male and female candidates depend on the position at stake. On the other hand, an additional female associate professor on the committee reduces the number of female candidates that are promoted by 7%. This effect reflects female associate professors' strong preference for male candidates from their own institution, relative to female colleagues. Replacing a male evaluator with a female evaluator from the same institution reduces the chances of success of female candidates from this institution by 38%, but it has no effect on the chances of success of female candidates from other institutions. The behavior of female associate professors may reflect strategic concerns. Perhaps they perceive that, because of an invisible gender quota, female assistant professors from their own institution may at some point compete with them for full professorship. The above results are not due to omitted characteristics of evaluators or candidates such as age, tenure, research production, research interests or academic and institutional links.

These findings have direct policy implications for countries and institutions that encourage female representation on hiring and promotion committees. Our results suggest that gender quotas in scientific committees will not necessarily increase female representation at the top. The effect of quotas will instead depend on how gendered academic networks are, as well as evaluators' strategic concerns, and the position at stake. Moreover, the benefits of quotas should be compared to their costs. Given the low numbers of women in top positions, quotas imply that women have to participate in committees much more often than men. Sitting on scientific committees reduces the available time for research; thus gender quotas might lower the productivity of women who have managed to overcome the glass ceiling.[6]

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. Finally, Section 6 concludes.

---

[6]Daniel Hamermesh warns young female economists to avoid requests to seat 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).

2 Institutional background

Before 2002, Spanish public universities had a large degree of autonomy regarding hiring and promotion. The system was characterized by a large degree of inbreeding: 93% of the positions were assigned to candidates from the same university offering the position (Cruz-Castro et al. 2006). In order to increase the transparency and meritocracy in the evaluation procedure, in 2002 the government introduced a system of centralized competitions known as *habilitación*. The system of *habilitación* is relatively similar to promotion systems currently in place in France and Italy.[7] This system requires candidates to associate and full professor positions to qualify in national competitions held at the discipline level.[8,9] Successful candidates can then apply for a position at a given university. In practice, during this period the number of vacancies created 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 *habilitación* system was in place until 2006, when it was replaced by a system known as *acreditación*, which is still in place.[10]

2.1 Centralized competitions between 2002 and 2006

The time line of examinations is as follows. First, the centralized competition is announced, and candidates are allowed to apply within the subsequent twenty days. Once

---

[7]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*).

[8]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.

[9]In total, there are nearly two hundred legally defined academic disciplines, each corresponding to a certain area 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). The list of disciplines is available at `http://www.educacion.gob.es/dctm/mepsyd/educacion/universidades/profesorado/habilitacion/areas-conocimiento.pdf?documentId=0901e72b80050626`.

[10]As in the *habilitación* system, in order to be eligible for promotion, candidates to associate and full professor positions are required to be accredited by a national evaluation committee. However, under the new system, committee members are not randomly drawn: they are selected by public officials from the pool of professors that volunteer for the task. Another relevant difference is that there is no limit to the number of candidates that may receive the accreditation. Since 2007, following the introduction of the Equality Law, evaluation committees also have to satisfy gender parity.

the list of applicants is settled, committee members are selected by random draw from the list of eligible evaluators. This list includes professors who have been officially recognized to have a minimum research quality in the discipline.[11] Each committee is composed of seven members. In exams to full professor positions, all committee members are chosen from the list of eligible full professors. In exams to associate professor positions, three committee members are selected from the set of eligible full professors, and four are selected from the set of eligible associate professors.[12] The committee member with the longest tenure is appointed president, and the exam is held at the university where the president is based. Evaluators can only resign under a very restricted set of reasons, and their resignation has to be officially approved.[13]

Exams to full professor positions have two qualifying stages. In the first stage, every candidate presents her *résumé*. In the second stage, candidates present a piece of their research work. Exams to associate professor, in addition to these two stages, have an intermediate stage where candidates give a lecture on a topic randomly chosen from a syllabus proposed by the candidate.

3   Data

We use data from three different sources. First, we collected data from all exams to associate and full professor positions that were held in Spain during the centralized system of examinations known as *habilitación* (years 2002 through 2006). Second, we gathered information on research output from Web of Science (ISI). According to survey information, Spanish professors consider publications in journals covered by ISI

---

[11]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 are required to hold at least one *sexenio*. Eligible full professors are required to hold at least two *sexenios*.

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

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

as the most important criterion in centralized examinations (Buela-Casal and Sierra 2006). Third, we exploit data on dissertations read in Spain since 1977. We use this information in order to identify the academic networks and the research interests of Spanish professors. In Appendix A we provide detailed information on how this information was collected.

3.1 Exams

Table 1 provides descriptive information on the characteristics of exams. In total there are 967 exams, of which 502 are exams to full professor positions and 465 are exams to associate professor positions. On average, in full professor exams there were three available positions per exam. In associate professor exams the figure was slightly higher; around five positions. The number of positions per candidate was very similar in both types of competitions; around 0.13-0.14 positions per candidate. Most vacancies were filled: 98% in exams to full professor positions, and 96% in exams to associate professor positions. The number of available positions per applicant is relatively similar across disciplinary areas. The disciplinary areas with the lowest ratios are Mathematics and Physics. In these disciplinary areas there were on average 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. In this case, there were 0.13 and 0.15 positions per candidate in exams to full and associate professor positions respectively.

3.2 Evaluators

Evaluators are randomly drawn out of the pool of eligible professors. Note that eligible professors are a selected sample of all professors. In order to be eligible as evaluators, professors are required to have a minimum level of research production. The minimum research level was satisfied by a relatively larger proportion of female professors than male professors. Around 80% and 84% of male and female full professors and approx-

imately 69% and 72% of male and female assistant professors respectively qualified.[14]

The upper panel of Table 2 provides information on the observable characteristics of eligible professors. There are 7,963 eligible full professors and 21,979 eligible associate professors. The average eligible full professor is 53 years old and has 13 years of tenure in this position. In terms of research, the average full professor has published seven papers in journals covered by ISI (we weight each publication by the inverse of the number of co-authors), received eight citations per paper, has supervised five dissertations, and has been invited 24 times to sit in a thesis committee. The latter measure may partly proxy for professional service. The average eligible associate professor is 45 years old and has 10 years of tenure. He/she has published four ISI articles, received seven citations per article, has supervised one thesis and has participated in five thesis committees.

The share of women decreases as we move up the career ladder. Women constitute 35% of associate professors, but only 14% of full professors are women. The presence of women also substantially differs across 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 most feminized disciplinary areas are Social Sciences (16%), Biology and Chemistry (17%), and the Humanities (18%).

Disciplines differ in many respects, including their propensity to publish and to cite. In our analysis we use information normalized to have zero mean and unit standard deviation for individuals within the same exam and category. This normalization also allows us to meaningfully compare female and male professors' characteristics (Table 2, columns (2) and (4)). Male full professors and male associate professors tend to be older and have longer tenure. They tend to publish more (7% of a standard deviation), but their publications are of similar quality, as measured by the number of citations received. Men supervise more theses and tend to participate in dissertation committees more often. Taking into account that the sample of eligible professors only includes

---

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

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, conditional on achieving this minimum, they have a slightly worse research output.

3.3 Candidates

Information on candidates is provided in the lower panel of Table 2. There are 13,601 applications to full professor positions, and 18,139 applications to associate professor positions. On average, candidates applied approximately twice during the period of study.

The average candidate to full professor positions is 46 years old, has four ISI publications and has received seven citations per publication. Compared to the average associate professor, candidates to full professor are slightly older and have a better research record. Not surprisingly, candidates to associate professor tend to be younger than candidates to full professors and have a relatively worse research record. On average they have only published two ISI articles, they have not advised any students and they have been invited to sit on a thesis committee only once.

There are relatively fewer female applicants in exams to full professor positions, the share of women among candidates to full and associate professor positions is 27% and 40% respectively. There are also gender differences in terms of the age of applicants. Male candidates to full professor positions tend to be younger than female candidates; however, male applicants to associate professor positions are older than their female counterparts. In full professor exams male candidates have accumulated a stronger research record: they have more publications, and they have supervised more students. In exams to associate professor positions, there are no significant gender differences in terms of the number of publications or the number of citations received by female-authored and male-authored work.

3.4 Links between evaluators and candidates

Evaluators and candidates may have some previous connection, which might potentially affect promotion decisions. Below we analyze both academic and institutional links.

3.4.1 Academic links

Within academic links, we distinguish between strong and weak academic ties.[15] We consider "strong academic ties" to be those situations where an evaluator and a candidate shared a previous academic connection. Following Zinovyeva (2011), we identify the following academic connections: (i) the evaluator was the candidate's dissertation advisor, (ii) the evaluator was a member of the candidate's thesis committee, (iii) the evaluator and the candidate had a common thesis advisor, (iv) the evaluator has invited the candidate to sit in one of her students' thesis committee (or vice versa), or (v) the evaluator and the candidate have a joint academic article. Some academic connections will not be captured by this taxonomy, but we believe this information to be informative of the extent and the characteristics of academic networks.

In exams to full professor positions, eligible evaluators and candidates have a 6.7% probability of being connected through any of these links (Table 3, columns (1)). The likelihood of a connexion is lower in exams to associate professor positions. In these exams, full professors in the committee have a 4.8% likelihood of being connected with candidates, for associate professors on the committee the probability of a link is 2.2% (Table 3, columns (4) and (7)).

There is an extensive literature that suggests that academic networks tend to be gendered. Hilmer and Hilmer (2007) observe that in the US 55% of the Economics PhD students being advised by women are female, while only 18% of Economics PhD students advised by men are female. 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. More informal networks might also be gendered. For instance, Mayer

---

[15]As Granovetter (1973) points out, the strength of a tie is a combination of the amount of time, the emotional intensity, the intimacy, and the reciprocal services which characterize the tie. We do not provide here an operational measure that allows to classify every academic tie.

and Puller (2008) use data from Facebook and observe that same gender is a strong predictor of friendship.

We examine if academic networks in Spain are gendered. In full professor exams, it is more frequent that male evaluators are linked to male candidates than to female candidates (7.0% vs 6.6%). On the contrary, female full professors are significantly more likely to be connected to a female candidate than a male candidate (6.0% vs 5.3%). In exams to associate professor positions, it is equally as likely for male evaluators to be linked to a male candidate and to a female candidate. However, the academic networks of female evaluators tend to be slightly gendered. Female full professors have links with 4.2% of male candidates and with 4.7% of female candidates to associate professor positions. Female associate professors are connected with 1.9% of male candidates and with 2.2% of female candidates. In sum, links are increasingly gendered both with the seniority of the candidate and the seniority of the evaluator. Moreover, we observe that evaluators, particularly women, are more likely to have links with same-sex candidates.

We consider "weak academic ties" to be those cases where we have not observed a strong academic link between an evaluator and a candidate, but it is reasonable to presume that they have interacted in the past. In particular, we assume that there exists a weak link between two individuals when their research interests are very close. We identify individuals' research interests at a very detailed level using the TESEO database of dissertations. In this database each thesis has been classified by its author using the Unesco International Standard Nomenclature for Fields of Science and Technology. This is a system developed by Unesco that includes more than two thousand categories. For instance, according to this classification, Economics is divided in one hundred different research fields. We take into account every dissertation where an individual has participated either as author, advisor, or committee member. Then, we use this information to measure how similar the research interests of candidates and evaluators are. We measure the *similarity* between two individuals i and j as the degree of overlap in their research interests:

$$Similarity_{i,j} = \sum_d \left(\frac{S_d^i + S_d^j}{2} \right) I(S_d^i > 0)I(S_d^j > 0)$$

where S_d^i is the ratio of the number of dissertations in field d in which individual i has participated over the total number of dissertations in which he/she has been involved, and I(.) denotes an indicator function. For instance, if one individual has participated in six dissertations classified as Microeconomic Theory and in four dissertations in Consumer Behavior, and another individual has participated in five dissertations in Microeconomic Theory and in five dissertations in Welfare Theory, the degree of similarity between these two individuals is equal to 0.55 ($= \frac{0.6+0.5}{2}$).[16]

In full professor exams, the degree of *similarity* between the research interests of candidates and evaluators is relatively large. In these exams candidates and evaluators on average share 36% of their research interests. The degree of similarity between candidates and full professor evaluators is 28%, and between candidates and associate professor evaluators it is 24% (Table 3, columns (2), (5) and (8)).

Several authors have observed that, within each discipline, men and women tend to specialize in different academic fields (for Economics, see Dolado et al. 2012, Hale and Regev 2011). Next, we investigate the existence of gender segregation across fields. We observe that, at every level, the research interests of male evaluators and female evaluators are equally close to the research interests of female candidates. However, male candidates tend to have a larger degree of similarity with male evaluators than with female evaluators.

3.4.2 Institutional links

Another potential source of connections is affiliation to a common institution. Next we examine how often evaluators and candidates come from the same institution and whether there is gender segregation across universities.

---

[16]Sometimes dissertations are assigned to several different fields. In this case we assign a proportional weight to each field.

In exams to full professor positions, we observe the university where evaluators and candidates are currently based. In exams to associate professor positions, we know evaluators' affiliation, but we cannot observe candidates' current affiliation. In this case we only have information on the university where they obtained their PhD. Given that in Spanish academia there is little institutional mobility, it might be reasonable to consider this information as a good proxy for current affiliation. To verify whether this is correct, we collected information from a random sample of 2,814 candidates.[17] We observe that 73% of them were based in the same university from which they graduated. In what follows we will use their PhD university as a proxy for the affiliation of candidates to associate professor positions.

In both full professor and associate professor exams there is approximately a 7% probability that an evaluator and a candidate are affiliated to the same academic institution. We do not observe any gender segregation in the upper rungs. In full professor exams, male and female evaluators are equally likely to be in the same institution as a given candidate (Table 3, column (2)). We find some evidence of gender segregation in exams to associate professor positions. In these exams female evaluators are more likely to be affiliated to the same university as female candidates (Table 3, columns (6) and (9)).

4   Empirical analysis

Our empirical analysis is structured as follows. First, we present our identification strategy, which exploits the random assignment of evaluators to committees. Second, we investigate whether the gender composition of academic committees affects applicants' chances of being promoted. Third, we discuss if the observed relationship between evaluators' gender and candidates' success reflects differences in evaluators' evaluations or differences in candidates' performance. Fourth, using information on applicants' research productivity, we analyze who is discriminating. Fifth, we test the

---

[17]Between April and June 2011, we searched online for information on the affiliation of 3,000 (randomly) chosen female candidates. We were able to find the whereabouts of 2,814 of them.

robustness of results to the inclusion of additional evaluators' characteristics. Finally, we explore the role played by the existence of prior academic and institutional links between evaluators and candidates.

4.1 Identification strategy

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, $E[s_e|\cdot] - \mu_e = 0$.

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.[18] 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. The expected proportion of female professors in the committee is equal to a weighted average of the proportion of women in each of the two pools.

The key assumption in our identification strategy is that the selection of committee members was 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 full professor exams, we observe that the expected proportion of female evaluators and the actual proportion of women

---

[18]As explained in footnote 12, 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 387 of the 967 exams in the sample. The details on these calculations are in Appendix B.

sitting on committees is statistically similar: 13-14%. Around 43% of (expected and actual) committees were composed by male evaluators only, and only in 3% of the cases there were 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 25%, approximately 19% of committees did not include any female members, 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 finally 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 by some other randomly chosen professors. 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.

4.2 Does the gender composition of committees matter?

We estimate the following linear probability model:[19]

$$y_{ie} = \beta_0 + \beta_1 f_i + \beta_2(s_e - \mu_e) + \beta_3 f_i(s_e - \mu_e) + \beta_4 z_e + \mathbf{d_f}\beta_5 + \epsilon_{ie} \tag{1}$$

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. To increase the accuracy of the estimation, we also control for the number of available positions per candidate (z_e) and for discipline fixed effects ($\mathbf{d_f}$). We cluster standard errors by exam 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.

We have rescaled z_e by subtracting its sample mean, and we consider the average

---

[19]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.

discipline as the reference group for discipline fixed effects. 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 of male applicants to full professor positions is 11.1%, about 1.7 percentage points higher than the success rate of female applicants. Since the specification does not include controls for candidate quality, this gender gap may reflect both the existence of gender differences in candidates' quality and discrimination by some evaluators. The gender gap in promotion is lower in committees with a relatively higher proportion of female evaluators. An additional female evaluator decreases the chances of success for male applicants by 6%, and increases the chances of success for female applicants by 14%.[20] Note that the relatively larger magnitude of the effect on female candidates reflects the lower proportion of women among candidates.

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: 12.6% of male candidates are promoted, compared to only 10.2% of female candidates. However, in this case an increase in the proportion of female full professors in the committee has no significant effect on the chances of success of male and female candidates (column 1). The effect of a larger presence of female full professors in exams to associate professor positions is significantly different from its effect in exams to full professor positions.[21]

---

[20] The average success of male candidates in full professor exams is 0.111. In a committee with seven members, an additional female evaluator decreases male candidates' chances of success by $(-0.047/7) * (1/0.111) * 100\% \approx -6\%$. The average success rate of female candidates is equal to 0.094, hence an additional female evaluator increases female candidates' chances of success by $[(-0.047 + 0.137)/7] * (1/0.094) * 100\% \approx 14\%$.

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

An increase in the number of associate professors in the committee negatively affects the chances of success for female applicants, though in this specification this effect is only significant at the 10% level. In quantitative terms, an additional female evaluator is associated with a 3% increase in the chances of promotion of male candidates, and a 7% decrease in the chances of promotion of female candidates.[22]

In sum, the gender composition of committees affects the outcome of promotion decisions, but the direction and the magnitude of the effect varies across the academic career ladder. 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, full professor evaluators do not have any gender preference. Moreover, a larger proportion of female associate professors among evaluators reduces the chances of promotion of female applicants.

4.3   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

---

[22]The average success of male candidates in associate professor exams is 0.126. In a committee with seven members, an additional female associate professor increases male candidates' chances of success by $(0.023/7) * (1/0.126) * 100\% \approx 3\%$. The average success rate of female candidates is equal to 0.102, hence an additional female evaluator decreases female candidates' chances of success by $[(0.023 - 0.071)/7] * (1/0.102) * 100\% \approx -7\%$.

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.[23] We divide candidates in three groups according to their distance to the university where the exam was held. There are no significant differences in the effect of committee gender composition across groups of candidates who face different travel costs (Figure 2). 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 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). We do not observe candidates' performance and thus we cannot completely discard the existence of a "stereotype threat". Nonetheless, the evidence does not offer support for this hypothesis. In theory, a potential "stereotype threat" would be more likely to occur in exams to associate professor positions than in exams to full professor positions. In the former, in addition to the two stages already present in full professor exams, candidates must give a lecture on one randomly chosen topic of the syllabus of an undergraduate course. Moreover, candidates are relatively less experienced and their performance might be more strongly affected by the attitude of the committee. 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 sum, the evidence is consistent with the idea that the gender composition of committees affects promotions directly through committee members' evaluations and it has little or no impact on candidates' performance.

---

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

4.4 Who is discriminating?

The identification strategy used above allows us to consistently estimate 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 estimate equation (1) taking into account candidates' quality.

We proxy for quality using ISI publications and received citations at the moment of applying to exam e. As a complementary measure of quality, we take into account the number of dissertations that the candidate has directed and the number of times he/she has participated in thesis committees. 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 these measures capture all gender differences in candidates' quality, we can analyze whether female candidates are being discriminated against by committees with a certain gender composition.

In exams to full professor positions, we do not observe any significant difference in the average success rate of male and female candidates once we control for candidates' (observable) quality (Table (5), column (2)). The average committee, which includes one female evaluator and six male evaluators, does not exhibit any gender bias. In exams to associate professor positions, male candidates have higher chances of success than comparable female candidates. In this case, the average committee, which includes two female evaluators and five male evaluators, seems to overrate male candidates.

The consistency of these estimates should be considered with caution. The identification strategy used here to identify the effect of candidates' gender relies on observable information on publication record and participation in dissertations. Women and men might differ in some relevant dimension other than research productivity. 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. One may also argue that, if promotion decisions are gender biased, a similar bias may be present in the publication process or in invitations for dissertations.[24]

4.4.1 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, $p(d_{je})$. Details about how these probabilities are calculated are provided in Appendix C. We estimate the following model, which allows for the gender composition of the committee to have a nonlinear effect on candidates' success rate:

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

---

[24]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 statistically significant at standard levels.

where d_{je} is a dummy variable that takes value one if the number of female evaluators in exam e is equal to j, and q_{ie} measures candidates' quality. Given the small number of committees with four, five or six female evaluators, we aggregate these committees into a single group.

In Figure 3 we plot the success of candidates to full professor positions (left figure) and associate professor positions (right figure) for different gender committee compositions.[25] In exams to both full professor and associate professor positions, 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.

We can also use this specification to examine the behavior of committees with a gender composition different from the average. In exams to full professor positions, in all-male committees female applicants 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 committees with female majority we observe a relatively large bias in favor of female candidates, but the difference is not statistically significant. We can only wonder if this effect would have been statistically significant if we had more committees with a female majority in exams to full professor positions.

In the case of exams to associate professor positions, women evaluated by all-male committees have the same chances of promotion as men. In committees with a majority of female evaluators, female candidates have significantly lower chances of success than male candidates of similar research quality.

4.4.2 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. Promotion decisions are then taken

---

[25]We report the results from estimating equation (2) in Appendix D, Table D1.

by the committee on a majority basis.

In exams to full professor positions, the evidence suggests that all-male committees discriminate against female candidates (based on the observable information on candidates' research production). The average committee, which includes one woman and six men, seems to treat male and female candidates equally. This evidence is consistent with male evaluators discriminating against female candidates when there are no women in the committee, but not doing so in the presence of female evaluators.

Results differ for exams to associate professor positions. If anything, our data suggests that, in all-male committees, 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.

4.5 Other committee characteristics

Men and women are different in many dimensions (other than gender) that could affect evaluations. Here we analyze if our results can be explained by differences in evaluators' age, tenure or research productivity.

Descriptive evidence presented in Table 2 suggests that male evaluators tend to be relatively older, have longer tenure, publish more, advise more students and participate in more thesis committees. We estimate equation (1) including the interaction between these characteristics and candidates' gender. 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 observe

23

any significant effect of these evaluators' characteristics on the chances of success of female and male applicants. The inclusion of these controls does not significantly affect our previous estimates (Table 5, column (3)).

4.6 The role of connections

Some authors have underlined the importance of networks for academic promotions (e.g. Combes et al. 2008). In a companion paper, Zinovyeva (2011) shows that, in Spanish academia, evaluators have a preference for candidates with whom they share some academic or institutional link. Below we study if connections mediate the effect of evaluators' gender on candidates' success. Additionally, we investigate if the effect of connections depends on the gender of candidates and evaluators.

4.6.1 Mediating the effect of connections

Connections tend to be gendered, particularly in full professor exams (see section 3.4). If evaluators favor their connections, and connections are of the same gender, this would result in same-sex candidates being overrated. We investigate if observable links between evaluators and candidates can explain our results.

First we examine to which extent evaluators' gender preferences can be explained by the existence of strong academic ties. We estimate equation (1) controlling for the links described in section 3.4.1. Accounting for academic networks does not have much effect on our previous estimates (Table 5, column (4)). In exams to full professor positions, the effect of the share of women in the committee on male candidates' chances of success decreases from 4.2 percentage points to 3.6 percentage points. The differential effect on female candidates decreases from 11.0 percentage points to 10.2 percentage points. We also investigate if academic ties play any role in explaining the gender patterns observed in associate professor exams. Accounting for these networks has no effect on our previous estimates.

Next, in order to assess to what extent our previous results reflect the existence of gender segregation across academic fields, we estimate equation (1) taking into account

similarity in the research interests of evaluators and candidates (Table 5, column (5)). We do not observe any noteworthy variation.

Finally, we consider institutional links. There exists gender segregation across universities in exams to associate professor positions, but not in exams to full professor positions (section 3.4.2). If men and women are based in different universities and evaluators have a preference for colleagues, this might generate a gender bias in promotion decisions. We investigate how our results vary if we consider affiliation, controlling for *same university* in equation (1). In full professor exams, affiliation has a very small effect on previous estimates (Table 5, column (6)). In associate professors exams, the effect of associate professors' gender on female candidates' chances of success becomes slightly larger in magnitude. The point estimate increases from 7.2 percentage points to 8.7 percentage points.

In sum, accounting for observable gender segregation across academic networks, research fields and universities has a limited effect on our previous results. In exams to full professor positions, the observed positive same-sex preference becomes slightly smaller. In the case of male candidates, this preference is not statistically significant anymore. In exams to associate professor positions, associate professors' preference for opposite-sex candidates becomes slighter larger.

4.6.2 Interaction between connections and gender

In order to get a better understanding of the role of connections, we explore how evaluators' behavior varies depending on the gender of the two individuals sharing a connection. That is, it might be that the way evaluators behave when they evaluate members in their own network depends not only on the existence of a link *per se*, but also on candidates' and evaluators' gender. As well, it might be that the strength of links depends on gender.

Full professor exams First we analyze the role of strong academic ties in exams to full professor positions and whether the gender of those involved matters. We estimate equation (1) taking into account the interaction between strong academic ties

and the gender of candidates and evaluators. As usual, we identify the causal effect of evaluators' characteristics by exploiting the difference between the expected number of evaluators with a certain characteristic and the actual realization. Candidates generally benefit from being evaluated by someone in their close network. This effect is larger if both the evaluator and the candidate are women. If a male evaluator is substituted by a female evaluator from the same academic network, chances of success of female candidates in the network increase by 79%, and the chances of success of female candidates that are not in the same network increase by 7%. This effect is statistically significant at the 5% level (Table 6, column (1)).

We also explore the interaction between research interests and the gender of candidates and evaluators. We classify the relationship between a candidate and an evaluator using a dichotomic variable which takes value one if they tend to do research in the same field. In particular, we consider that an evaluator and a candidate are in the same field of research if they share at least 50% of their research interests (*similarity* $\geq$ *0.5*).[26] Most interestingly, evaluators do not exhibit any gender preference when they evaluate candidates whose research is in a different field (Table 6, column (2)). Success depends on gender only when the evaluators and the candidate share the same research interests. In this case, same-sex candidates tend to be favored. The magnitude of the effect is relatively large. If we replace a male evaluator by a female evaluator from the same research field, the chances of success of female candidates in this field increase by 1.7 percentage points (17%). This effect is significant at the 5% level. At the same time, the chances of success of male candidates in this field decrease by 1.4 percentage points (13%), even though in this case this difference is not significant at standard levels.[27]

---

[26]Approximately in 55% of cases the degree of similarity between a candidate and an evaluator is equal to zero, in 25% of cases it is between zero and 50%, and in 20% of cases it is above 50%. We obtain qualitatively similar results if we reduce the threshold to any value between zero and 50%.

[27]In the average committee the chances of success of female candidates are equal to 0.101 (0.109-0.008). Female candidates' chances of success increase in 1.71 percentage points [((0.159+0.037)-(0.138-.062)/7)] if we replace a male evaluator from the same field as the candidate by a female evaluator from that field. This is equivalent to a 17% increase (1.71/10.1). Analogously, in the average committee the success rate of male candidates is equal to 0.109. Male candidates' chances of success decrease in 1.4 percentage points ((0.138-0.037)/7), or 13% (0.014/0.109).

Finally, we look at the interaction of gender and *same university*. Candidates benefit from having a senior colleague in the evaluation committee, but gender is not relevant. If we replace a male evaluator with a female evaluator from the same institution, this has no significant effect on the chances of success of male and female candidates from that institution (but it increases the chances of success of female candidates from other institutions). To sum up, in full professor exams, gender only matters when evaluators and candidates are in the same network, or whenever their research is in the same field, even if they are not affiliated to the same institution.

Associate professor exams We proceed in a similar way in exams to associate professor positions. As we observed earlier (Table 5), in these exams female candidates tend to have lower possibilities of success than equally qualified male candidates.

This gap is lower when candidates are evaluated by a female full professor from the same institution as themselves, relatively to a male full professor from the same institution, but this difference is not statistically significant (p-value=0.14). On the contrary, the gap is significantly larger if candidates are evaluated by a female associate professor from their institution, relatively to a male colleague (Table 6, column (6)). The magnitude of the effect is statistically and economically significant. Replacing a male associate professor with a female associate professor from the same institution reduces the chances of success of female candidates from that institution in 3.91 percentage points (38%), and has no effect on female candidates from other institutions.[28] In sum, female associate professor evaluators exhibit an opposite-sex preference only when they assess the quality of their colleagues.

Note that, unlike in full professor exams, in associate professor exams the effect of academic networks and research interests is not gendered. Female evaluators tend to have a preference for male candidates from their own institution, but this does not depend on the existence of academic links or similar research interests (Table 6,

---

[28]In the average committee the rate of success of female candidates to associate positions is equal to 0.103 (0.126-0.023). If we replace a male associate professor with a female associate professor from the same institution, female candidates from that institution experience a 3.91 percentage points decrease in their chances of success [((0.441-0.433)-(0.325-0.043))/7]. This is equivalent to a 38% reduction (3.91/10.3).

columns (4) and (5)).

5   Discussion

The evidence suggests that, in exams to full professor positions, evaluators favor same-sex candidates whose research is in the same field, particularly if they are in the same academic network. In exams to associate professor positions, both male and female evaluators tend to prefer male candidates. The gender gap is larger, though, when candidates are evaluated by a female associate professor from their own institution. Below, we discuss several theories that might explain these findings.

5.1   Full professor exams

There are two set of explanations that can potentially explain these results: information asymmetries and taste discrimination.

5.1.1   Information asymmetries

In academia, evaluating ability often requires possessing some related knowledge. Otherwise, it is not possible to accurately assess the quality of candidates. In the presence of such information asymmetries, evaluators may be affected by gender stereotypes. For instance, some men may hold the prejudice that women tend to be worse researchers. These stereotypes are expected to be more relevant when evaluators are not familiar with the candidate's research field. However, we observe the opposite. Candidates' gender only matters when evaluators have research in the same field. This suggests that gender stereotypes are not the source of the pattern observed in exams to full professor positions.

Information asymmetries can also cause gender discrimination even if evaluators do not hold gender prejudices or gender stereotypes (Cornell and Welch 1996, Bagues and Perez-Villadoniga 2008). The variance of candidates' inferred quality tends to be relatively smaller if the evaluator cannot accurately assess candidates' true quality.

28

As a result, when promotion decisions are decided through a tournament where only the best candidates are selected, the candidates that the evaluator can assess more accurately tend to have better chances of being promoted. If evaluators were able to better assess the quality of same-sex candidates who do research in the same field, that would explain the pattern we observe. However, there is no reason to believe that, within a given field, evaluators' accuracy depends on candidates' gender.

5.1.2 Taste discrimination

Our results might also reflect the existence of taste discrimination. There are several potential sources for this. Evaluators may generally dislike candidates from the opposite gender. However, this is at odds with the empirical evidence. Evaluators only take into account the gender of candidates with research in the same field as themselves.

Gender discrimination could also arise as a consequence of cronyism. If evaluators tend to favor friends and friendships tend to be gendered, this would result in same-sex candidates being favored. However, taking account observed links between candidates and evaluators explains only a very limited part of the gender effects (Table 5).

Our measure of links may be subject to measurement error, as it does not take intensity into account. It might be that links are stronger for same-sex individuals. In fact, we find that full professors favor only same-sex candidates with research in the same field, defined at a very detailed level. Moreover, the preference for same-sex candidates is largest when evaluators have to decide on the promotion of candidates in their close academic network. This evidence is consistent with evaluators being more attached to same-sex connections. All in all, our results are consistent with promotions to full professor being affected by the existence of *old boys'* and *old girls'* networks.

5.2 Associate professor exams

In exams to associate professor positions, the gender of full professor evaluators is not relevant. The gender of associate professor evaluators only affects the chances of promotion of candidates in their own institution.

5.2.1 Full professor evaluators

The difference between the behavior of full professors in exams to full professor positions and their behavior in exams to associate professor positions is statistically significant at the 5% level. We might hypothesize several possible explanations for this differential conduct. First, it might be that full professors have weaker links with same-sex candidates to associate professor positions. This is consistent with the fact that in associate professor exams we observe fewer links between candidates and evaluators (Table 3). This might be related to differences in age: candidates to full professor positions are on average six years younger than evaluators. In exams to associate professors this difference is fifteen years.

An alternative interpretation of the evidence is that full professors may have different gender attitudes depending on the position at stake. According to the theory of ambivalent sexism, sexism might be a "multidimensional construct that encompasses two sets of sexist attitudes" (Glick and Fiske 1996). Male evaluators might experience sexist antipathy towards female candidates applying to top academic positions, but subjectively feel indifferent (or even a positive orientation) toward female candidates applying to lower-level positions.

5.2.2 Associate professor evaluators

In exams to associate professor positions, female candidates are less likely to be promoted when they are evaluated by a committee that includes a female associate professor from their institution, relative to a male colleague. There are several theories that might explain female evaluators' preference of opposite-sex colleagues.

It might be that, as in Bagues and Esteve-Volart (2010), female evaluators overestimate the quality of male candidates.[29] Further, in a framework of information asymmetries, female evaluators may decide to be more demanding with female candidates in order to increase the average quality of female professors. This would positively

---

[29]Bagues and Esteve-Volart (2010) analyze entry exams to the Spanish Judiciary. Evaluators can only base their decision on candidates' performance in several tests where they are evaluated on their (memoristic) knowledge on a number of topics. In total, the evaluation process takes two or three hours. In a set up like that, information asymmetries may be very important.

affect the priors that are commonly held about the quality of female professors, including evaluators. A prediction of this theory is that female evaluators should be relatively tougher with female applicants in less feminized fields. However, the data is not consistent with this implication. When we examine results separately for several fields (ordered by their degree of feminization) we do not observe any trend (Figure 4). This result also rules out the *self-enhancement drive* theory, which argues that women in male-dominated fields tend to identify with male rather than female colleagues, in order to maintain a positive social identity (Graves and Powell 1995).

Associate professors still have to compete to be promoted to full professor positions. In Spanish universities full professors tend to be selected out of the pool of associate professors at the institution. Female associate professors might perceive female candidates as potential competitors if, for some reason, they think that the number of women that will be promoted to full professor positions in the future is limited by some (invisible) gender quota. This theory would explain why female associate professor evaluators discriminate against their female colleagues, but not against other female candidates (and why, simultaneously, female full professors do not have any gender preference in these exams). Alternatively, our results might reflect some sort of non-academic rivalry among female junior faculty (Ruffle and Shtudiner 2010)

6 Conclusions

In the last few decades there has been a significant increase in the number of women starting academic careers both in the US and in Europe. However, 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.

This paper studies whether promotion committees hinder women's access to top positions. 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 in all academic disciplines. These competitions involved around 30,000 candidacies and 7,000 evalua-

tors. 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 its effect depends on a number of factors. In full professor exams, evaluators tend to favor same-sex candidates with whom they share some academic link. This pattern is consistent with the existence of *old boys* and *old girls* networks. On the other hand, in associate professor exams, full professor evaluators do not exhibit any gender preference, but female associate professor evaluators discriminate against female candidates from their institution, perhaps for strategic reasons. The differential behavior of full professors may reflect the lesser importance of networks at the early stages of the academic career. Alternatively, it might be that evaluators' attitudes toward male and female candidates depend on the position at stake. Associate professors' preference for opposite-sex colleagues might be due to strategic concerns. It may be that female associate professors expect full professor positions to be subject to an invisible gender quota. These results are not due to omitted characteristics of evaluators such as age, tenure, research production, research interests, affiliation or academic connections.

Our findings have important policy implications. To prevent gender discrimination, several countries, including Norway (1988), Finland (1995), Sweden (1999) and Spain (2007), have introduced gender quotas in scientific committees (European Commission 2008). Since, as shown in this paper, quotas may have a positive effect on female promotion only in certain situations, our work provides strong evidence against a generalized implementation of gender quotas. The suitability of quotas will depend on the degree to which networks are gendered, evaluators' strategic concerns, and the position at stake.

In the case of Spain, gender quotas exist in all hiring and promotion committees in public institutions, including universities. The quota mandates that at least 40% of committee members must be of each gender. According to our findings, the introduction of a 40% gender quota in full professor committees is excessive. In this

case, the presence of a single female on the committee would be enough to overcome the bias induced by *old boys* networks. On the contrary, at the lower rungs of the academic ladder quotas decrease the number of women who are promoted. Moreover, quotas increase disproportionally the time that female professors have to spend sitting on committees.[30]

References

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.

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

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).

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

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

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.

Blau, Francine, Marianne Ferber and Anne Winkler (2010), *The Economics of Women, Men and Work*, 6th ed. Upper Saddle River, NJ: Pearson/Prentice-Hall.

Booth, Alison L. and Andrew Leigh (2010), "Do Employers Discriminate by Gender? A Field Experiment in Female-Dominated Occupations," *Economic Letters*, Vol. 107(2), pp. 236-238.

Boschini, Anne and Anna Sjögren (2007), "Is Team Formation Gender Neutral? Evidence from Coauthorship Patterns," *Journal of Labor Economics*, Vol. 25(2), pp. 325-365.

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

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.

---

[30]A back-of-the-envelope calculation suggests that a female full professor has to sit in committees four times more often than a male full professor. 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 $[(0.85 - (0.85 * 0.60))/(0.15 * 0.60) \approx 4]$.

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.

Cornell, Bradford and Ivo Welch (1996) "Culture, Information, and Screening Discrimination," *Journal of Political Economy* , Vol. 104(3), pp. 542-571.

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.

Dolado, Juan Jose, Florentino Felgueroso and Miguel Almunia (forthcoming), "Are men and women-economists evenly distributed across research fields? Some new empirical evidence," *SERIEs*.

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

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

Fuentes, Eulàlia and Llorenç Arguimbau (2010), "Las tesis doctorales en España (1997-2008): análisis, estadísticas y repositorios cooperativos," *Revista Española de Documentación Científica*, Vol. 33(1), pp. 63-89.

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.

Ginther, Donna K. and Shulamit Kahn (2004), "Women in Economics: Moving Up or Falling Off the Academic Ladder," *Journal of Economic Perspectives*, Vol. 18(3), pp 193-214.

Ginther, Donna K. and Shulamit Kahn (2009), "Does Science Promote Women? Evidence from Academia 1973-2001," NBER chapters in: *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, pp. 163-194.

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.

Granovetter, Mark S. (1973), "The Strength of Weak Ties," *American Journal of Sociology*, Vol. 78(6), pp. 1360-1380.

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.

Hale, Galina and Tali Regev (2011), "Gender Ratios at Top PhD Programs in Economics," mimeo.

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

Hilmer, Christiana and Michael Hilmer, "Women Helping Women, Men Helping Women? Same-Gender Mentoring, Initial Job Placements, and Early Career

Publishing Success for Economics PhDs," *American Economic Review*, Vol. 97(2), pp. 422-426.

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

Instituto Nacional de Estadística, Estadística de la Enseñanza Universitaria, several issues.

National Science Foundation, Division of Science Resources Statistics (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/*.

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

National Research Council (2009), *Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty*, Washington D.C.: The National Academies Press.

Mayer, Adalbert and Steven L. Puller (2008), "The old boy (and girl) network: Social network formation on university campuses," *Journal of Public Economics*, Vol. 92(1-2), pp. 329-347.

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*, Vol. 89(2), pp. 392-396.

Ruffle, Bradley J. and Ze'ev Shtudiner (2010), "Are Good-Looking People More Employable?," *Ben-Gurion University of the Negev Discussion Paper* 10-06

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

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.

Wennerås, Christine and Agnes Wold (1997), "Nepotism and sexism in peer-review," *Nature*, Vol 387, pp. 341-343.

Zinovyeva, Natalia (2011), "It's not what you know, but who you know: The role of connections in academic promotion," *mimeo*, IPP-CSIC.

Table 1: Descriptive statistics − Examinations

	Full-professor exams (N=502)				Associate-professor exams (N=465)			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Positions per exam	2.92	1.78	1	12	4.74	4.71	1	25
Candidates per exam	27.09	17.98	3	132	39.01	34.82	3	270
Positions per candidate	0.13	0.07	0.02	0.67	0.14	0.08	0.02	0.67
Share of positions filled	0.98	0.09	0	1	0.96	0.15	0	1

Table 2: Descriptive statistics − Eligible evaluators and candidacies

	1	2	4	5
	Eligible Evaluators			
	Full professors		Associate professors	
	Total	Difference between men and women, normalized at the exam level	Total	Difference between men and women, normalized at the exam level
Female	0.14	-	0.35	-
	(0.35)		(0.48)	
Age	52.90	0.08***	44.98	0.04***
	(6.41)	[0.01]	(7.82)	[0.01]
Tenure in position	12.94	0.39***	10.36	0.04***
	(8.20)	[0.01]	(6.59)	[0.01]
Publications, weighted by co-authors	7.24	0.07***	3.51	0.04***
	(11.03)	[0.01]	(5.05)	[0.01]
Citations per publication	8.07	-0.01	7.30	0.01
	(9.74)	[0.01]	(9.87)	[0.01]
PhD students advised	4.85	0.24***	1.09	0.22***
	(5.07)	[0.01]	(1.98)	[0.01]
PhD committees	23.94	0.38***	4.65	0.20***
	(23.71)	[0.01]	(7.00)	[0.01]
Total number of observations	49,199		61,052	
Total number of individuals	7,963		21,979	
	Candidacies			
	Full professor exams		Associate professor exams	
	Total	Difference between men and women, normalized at the exam level	Total	Difference between men and women, normalized at the exam level
Female	0.27	-	0.40	-
	(0.44)		(0.49)	
Age	46.40	-0.05***	37.46	0.08***
	(6.50)	[0.02]	(6.55)	[0.02]
Publications, weighted by co-authors	4.35	0.07***	2.33	-0.02
	(6.08)	[0.02]	(3.90)	[0.01]
Citations per publication	7.38	-0.01	6.18	-0.01
	(10.12)	[0.02]	(12.16)	[0.01]
PhD students advised	1.76	0.12***	0.20	0.07***
	(2.52)	[0.02]	(0.81)	[0.01]
PhD committees	6.69	0.11***	0.73	0.13***
	(8.26)	[0.02]	(2.39)	[0.01]
Total number of observations	13,601		18,139	
Total number of individuals	6,539		10,039	

Notes: Mean values, standard deviations in parentheses, standard errors in square brackets. 547 full professor eligible evaluators and 645 associate professor eligible evaluators have missing information on age. 37 full professor eligible evaluators and 249 associate professor eligible evaluators have missing information on the length of tenure. In these cases we imputed the missing information using the mean value in the corresponding exam and rank. 457 candidates to full professor positions and 396 candidates to associate professor positions have missing information on age. Citations per publication are equal to zero when an individual has not published.

Table 3: Gender pattern of links between candidates and eligible evaluators

	1	2	3	4	5	6	7	8	9
	Candidates to full professor* Full professor evaluators			Candidates to associate professor* Full professor evaluators			Candidates to associate professor* Associate professor evaluators		
	Academic network	Field similarity	Same university	Academic network	Field similarity	Same university	Academic network	Field similarity	Same university
Female candidate	-0.004**	0.006	0.005***	-0.001	0.002	-0.001	-0.000	-0.000	-0.001
	(0.002)	(0.004)	(0.002)	(0.001)	(0.004)	(0.002)	(0.001)	(0.003)	(0.002)
Female evaluator	-0.017***	-0.013***	0.001	-0.007***	-0.008***	-0.001	-0.003***	-0.006***	0.002*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Female candidate*Female Evaluator	0.011***	0.012***	-0.001	0.006***	0.006	0.007**	0.003***	0.007***	0.007***
	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)	(0.002)
Constant	0.070***	0.358***	0.065***	0.049***	0.276***	0.070***	0.022***	0.239***	0.061***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.002)	(0.001)
Mean of the dependent variable	0.067	0.359	0.067	0.048	0.277	0.070	0.022	0.238	0.062
Number of observations	490220	476874	490220	757663	625959	623678	2195186	1812760	1812955

Notes: OLS estimates, all candidacies are given the same weight. The number of observations is the number of possible matches between candidacies and eligible evaluators with non-missing information. Standard errors clustered by candidate are in parentheses. All regressions include discipline fixed effects. In columns 1, 4, and 7 the dependent variable is a dummy variable that takes value one if there exists some academic link between candidate and evaluator. In columns 2, 5, and 8 the dependent variable is a continuous variable that measures the overlap in research interests. In column 3 the dependent variable is a dummy that takes value one if candidate and evaluator are affiliated to the same university. In columns 6 and 9 the dependent variable is a dummy that takes value one if the candidate graduated in the same university where the evaluator is based.

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	0.13	0.14	0.25	0.25
	(0.00)	(0.01)	(0.01)	(0.01)
Zero female evaluators	0.43	0.43	0.20	0.19
	(0.01)	(0.02)	(0.01)	(0.02)
One female evaluator	0.32	0.31	0.27	0.27
	(0.01)	(0.02)	(0.01)	(0.02)
Two female evaluators	0.16	0.16	0.25	0.27
	(0.01)	(0.01)	(0.01)	(0.02)
Three female evaluators	0.06	0.07	0.16	0.17
	(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	53.16	53.19	48.76	48.95
	(0.11)	(0.15)	(0.14)	(0.18)
Tenure in position	13.31	13.27	11.69	11.73
	(0.10)	(0.16)	(0.10)	(0.15)
Publications, weighted by co-authors	5.76	5.43	3.79	3.81
	(0.27)	(0.27)	(0.18)	(0.21)
Citations per publication	7.01	6.91	6.20	6.18
	(0.25)	(0.27)	(0.24)	(0.27)
PhD students advised	4.90	4.83	2.80	2.87
	(0.10)	(0.12)	(0.07)	(0.09)
Membership in PhD committees	23.77	24.36	12.75	12.90
	(0.43)	(0.55)	(0.27)	(0.34)

Notes: The sample includes 502 exams to full professor positions and 465 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
	Full professor exams					
Female candidate	-0.017***	-0.010	-0.010	-0.010	-0.009	-0.009
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Female FP evaluators	-0.047***	-0.046***	-0.042**	-0.036**	-0.031*	-0.029
	(0.017)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)
Female candidate*Female FP evaluators	0.137***	0.123**	0.110**	0.102**	0.105**	0.102**
	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
Constant	0.111***	0.110***	0.110***	0.110***	0.110***	0.110***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Candidates' characteristics		✓	✓	✓	✓	✓
Evaluators' characteristics			✓	✓	✓	✓
Academic networks				✓	✓	✓
Research field similarity					✓	✓
Same university						✓
Adjusted R-squared	0.022	0.052	0.051	0.063	0.064	0.066
Number of observations	13601	12680	12680	12680	12371	12371
	Associate professor exams					
Female candidate	-0.024***	-0.023***	-0.022***	-0.022***	-0.022***	-0.021***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Female FP evaluators	0.012	0.015	0.029	0.033	0.047	0.054
	(0.030)	(0.030)	(0.032)	(0.031)	(0.038)	(0.043)
Female candidate*Female FP evaluators	-0.013	-0.014	-0.044	-0.040	-0.038	-0.041
	(0.064)	(0.063)	(0.065)	(0.065)	(0.075)	(0.077)
Female AP evaluators	0.023	0.029*	0.033**	0.027*	0.014	0.028
	(0.016)	(0.016)	(0.016)	(0.016)	(0.020)	(0.022)
Female candidate*Female AP evaluators	-0.071*	-0.069*	-0.073**	-0.073**	-0.072*	-0.087*
	(0.037)	(0.036)	(0.037)	(0.037)	(0.043)	(0.044)
Constant	0.126***	0.126***	0.125***	0.125***	0.132***	0.129***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Candidates' characteristics		✓	✓	✓	✓	✓
Evaluators' characteristics			✓	✓	✓	✓
Academic networks				✓	✓	✓
Research field similarity					✓	✓
Same university						✓
Adjusted R-squared	0.026	0.038	0.037	0.050	0.053	0.059
Number of observations	18139	17432	17432	17432	14789	13213

Notes: OLS estimates. Standard errors are clustered by exam. All regressions include discipline fixed effects and the number of positions per candidate. These controls are rescaled to have zero mean in the corresponding subsample, therefore the constant is the average success rate of male candidates while the female dummy is the gender gap in the average committee. FP and AP stand for full and associate professor respectively. Female FP/AP evaluator is the difference between the actual and the expected proportion of female evaluators (of a given rank) among all committee members. Candidates' characteristics include age, publications, citations per publication, number of students advised and the membership in PhD committees. All these controls are normalized at the exam level. Columns (3)-(6) include the difference between the actual and the expected committee composition in terms of mean age, tenure, number of publications, citations per publication, number of students advised and the membership in PhD committees, and the corresponding interactions with the indicator for female candidates. Columns (4)-(6) include the difference between the actual and the expected academic network proximity between candidates and evaluators. Columns (5)-(6) include the difference between the actual and the expected overlap in research interests between candidates and evaluators. Column (6) includes the difference between the actual and the expected proportion of colleagues in the committee.

Table 6: When does gender matter?

|	1	2	3	4	5	6
	Full professor exams			Associate professor exams		
	X is equal to:			X is equal to:		
	Academic network	Field of research	Affiliation	Academic network	Field of research	Affiliation
Female candidate	-0.009	-0.008	-0.010	-0.024***	-0.022***	-0.023***
	(0.006)	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)
Female FP evaluators, $X_{cand} \neq X_{eval}$	-0.034*	-0.010	-0.044**	0.023	0.034	0.036
	(0.018)	(0.026)	(0.019)	(0.031)	(0.054)	(0.039)
Female candidate*Female FP evaluators, $X_{cand} \neq X_{eval}$	0.085*	0.073	0.115**	-0.007	-0.010	-0.047
	(0.050)	(0.064)	(0.052)	(0.064)	(0.090)	(0.067)
Male FP evaluators, $X_{cand} = X_{eval}$	0.448***	0.138***	0.343***	0.829***	0.205***	0.466***
	(0.057)	(0.030)	(0.047)	(0.111)	(0.057)	(0.079)
Female candidate*Male FP evaluators, $X_{cand} = X_{eval}$	-0.093	-0.062	-0.072	-0.226	-0.084	0.038
	(0.100)	(0.057)	(0.084)	(0.165)	(0.087)	(0.115)
Female FP evaluators, $X_{cand} = X_{eval}$	0.300	0.037	0.372**	0.751**	0.206**	0.149
	(0.184)	(0.045)	(0.146)	(0.316)	(0.100)	(0.223)
Female candidate*Female FP evaluators, $X_{cand} = X_{eval}$	0.609*	0.159*	0.030	-0.319	-0.037	0.755**
	(0.325)	(0.091)	(0.229)	(0.428)	(0.143)	(0.331)
Female AP evaluators, $X_{cand} \neq X_{eval}$				0.030*	0.035	0.043**
				(0.016)	(0.026)	(0.021)
Female candidate*Female AP evaluators, $X_{cand} \neq X_{eval}$				-0.077**	-0.099*	-0.060
				(0.036)	(0.054)	(0.042)
Male AP evaluators, $X_{cand} = X_{eval}$				0.609***	0.090*	0.325***
				(0.142)	(0.050)	(0.078)
Female candidate*Male AP evaluators, $X_{cand} = X_{eval}$				-0.358	-0.131*	-0.043
				(0.245)	(0.074)	(0.123)
Female AP evaluators, $X_{cand} = X_{eval}$				0.540*	0.066	0.441***
				(0.307)	(0.065)	(0.121)
Female candidate*Female AP evaluators, $X_{cand} = X_{eval}$				-0.066	-0.161*	-0.433***
				(0.424)	(0.096)	(0.161)
Constant	0.110***	0.109***	0.110***	0.126***	0.130***	0.126***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Candidates' characteristics	✓	✓	✓	✓	✓	✓
Adjusted R-squared	0.063	0.053	0.059	0.048	0.042	0.049
Number of observations	12680	12371	12680	17432	13213	14911

Notes: OLS estimates. Standard errors are clustered by exam. Candidates' characteristics include age, publications, citations per publication, number of students advised and the membership in PhD committees. All these controls are normalized at the exam level. All regressions include discipline fixed effects and the number of positions per candidate. These controls are rescaled to have zero mean in the corresponding subsample, therefore the constant is the average success rate of male candidates while the female dummy is the gender gap in the average committee. FP and AP stand for full and associate professor respectively. Female FP/AP evaluator is the difference between the actual and the expected proportion of female evaluators (of a given rank) among all committee members. 'Same field of research' indicates whether a candidate and an evaluator have the overlap of research interests exceeding 50%. In columns (1)-(3), the reference group is a 'Male FP evaluator, $X_{cand} \neq X_{eval}$'. In columns (4)-(6), the reference groups are 'Male FP evaluators, $X_{cand} \neq X_{eval}$' and 'Male AP evaluators, $X_{cand} \neq X_{eval}$'.

**Figure 1:** Proportion of women in Spanish academia
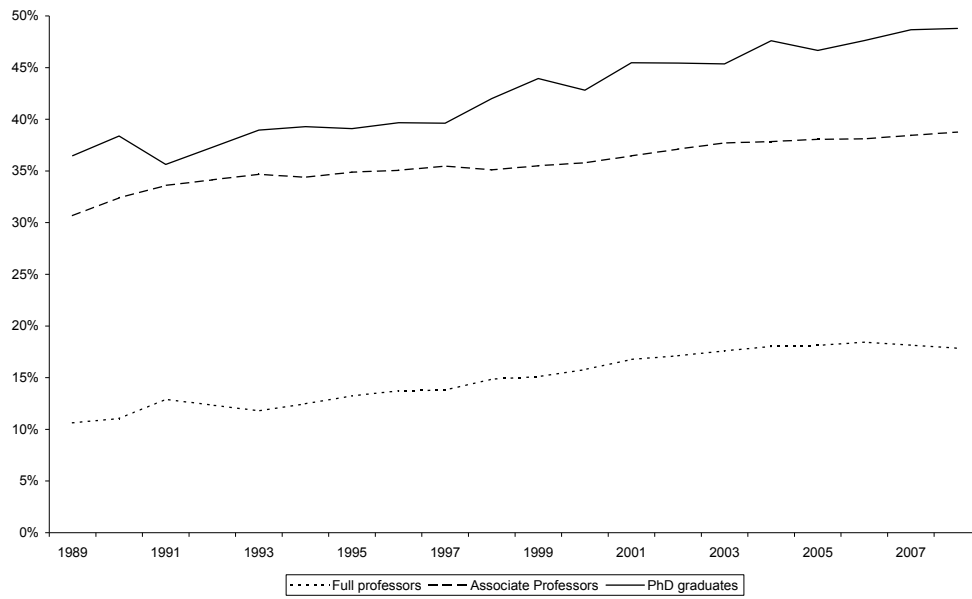
Source: Authors' calculations based on information from Instituto Nacional de Estadística, Estadística de la Enseñanza Universitaria, several issues.

Figure 2: The effect of committees' gender composition on the relative advantage of female candidates, by travel time
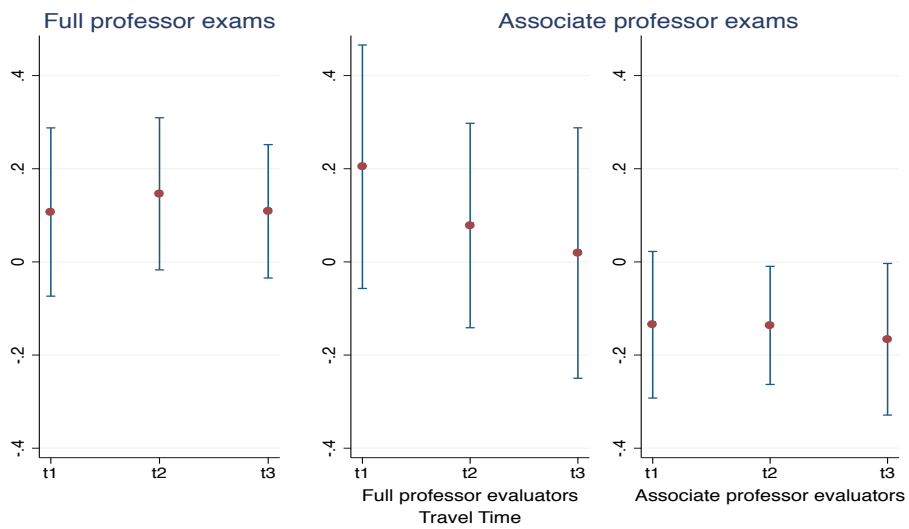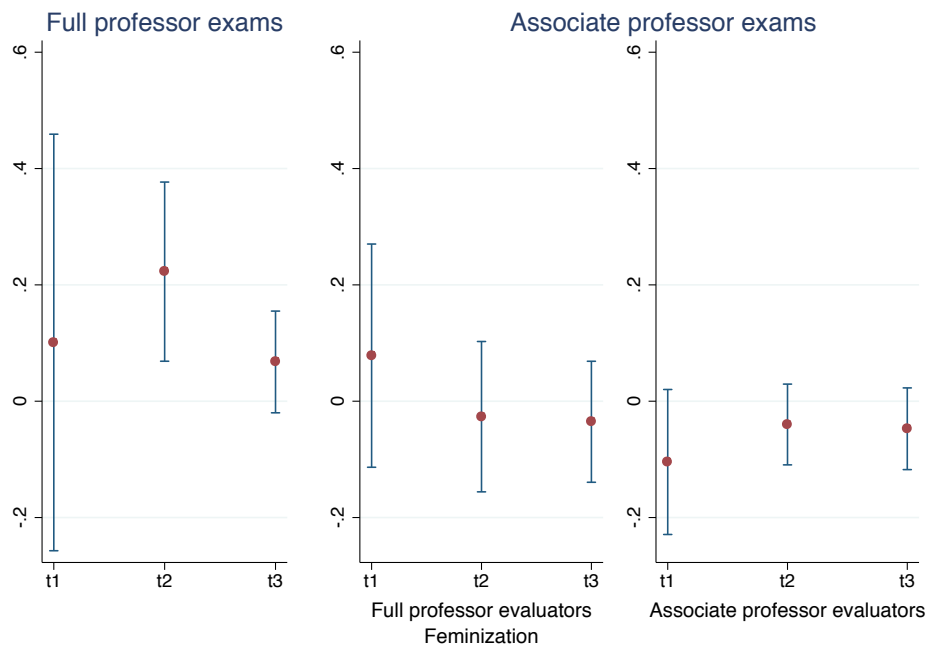

Notes: The y-axis represents the effect of gender composition on the gender gap in promotions, as estimated in Table D2, columns (1) and (3). In the x-axis candidates are grouped according to the distance between their home university and the location where the exam takes place. Group *t1* includes candidates in the first tercile of the distribution in terms of distance. These candidates are based in a university which is approximately less than four hours away by car from the university where the exam took place. Candidates in the second tercile (*t2*) are based in a university located between four and six hours away. Candidates in the third tercile (*t3*) work in universities more than six hours away. 95% confidence intervals are displayed.

Figure 3: The effect of committees' gender composition on the success rate of male and female candidates
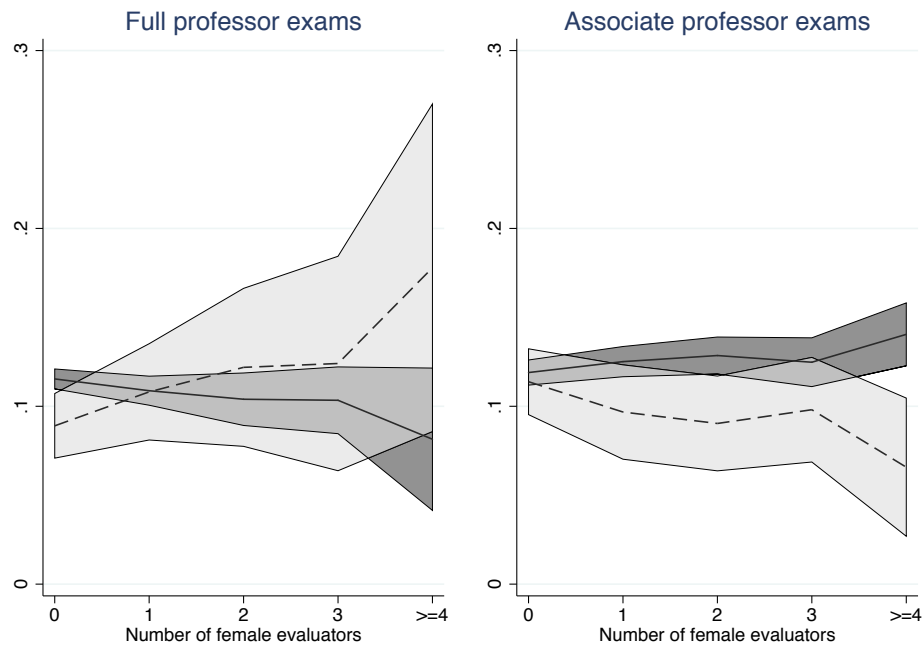
Notes: The solid (dashed) line represents the success rate of male (female) candidates, as estimated in Table D1. Committees are composed of seven members. 95% confidence intervals are displayed.

Figure 4: The effect of committees' gender composition on the relative advantage of female candidates, by degree of feminization of the field

Notes: The y-axis represents the effect of gender composition on the gender gap in promotions, as estimated in Table D2, columns (2) and (4). In the x-axis candidates are grouped according to the degree of feminization of their discipline. Group *t1* includes candidates in the first tercile of the distribution in terms of the degree of feminization of the field. In these disciplines approximately less than 20% of tenured professors are women. In disciplines in the second tercile women constitute between 20% and 35% of tenured professors. In disciplines in the third tercile at least 35% of tenure professors are women. 95% confidence intervals are displayed.

Appendix A: Data Appendix

We have collected information from three different sources: (i) Ministry of Research and Science, (ii) ISI Web of Science and (iii) Teseo database on doctoral dissertations. Below we describe the process of data collection in detail.

Ministry of Research and Science The system of centralized examinations known as 'habilitación') was in place between 2002 and 2006. 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`). In total, 1,016 exams took place, around five per discipline. We restrict the sample in several ways. We exclude exams where the number of available positions was larger or equal than the number of candidates (two exams, both in Basque Philology) and disciplines where the number of potential evaluators was not large enough to form a committee (55 exams).[31] The final database includes 967 exams.

We used first name information in order to identify gender. In a few cases where it was not possible to assign gender based on first name, we searched online for any personal picture or document that would made it possible to assign gender.

The actual age of individuals is not observable. Instead, we exploit the fact that Spanish ID numbers contain information on their issue date to construct a proxy for the age of native individuals on the basis of his/her national ID number. In Spain, police stations are given a range of ID numbers, which they 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).[32] 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

---

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

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

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 the cases the assigned age is within a three-year interval of the actual age. In order to minimize potential errors, whenever our age proxy indicated that a candidate to an associate professor position is less than 27 years old and a candidate to full professor positions 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 assume that eligible evaluators for whom the age proxy is missing, are 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).

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

ISI Web of Science   We have also collected information on the research output of eligible evaluators and candidates from the ISI Web of Science.[33]

Information on scientific publications comes from the 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. The 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

---

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

conferences. For the purpose of this analysis, we considered all articles, reviews, notes and proceedings.

The assignment of articles to professors is non trivial. For each publication and author, WoS provides information on his/her surname and on his/her initial. In Spain, some surnames are very common (e.g., Garcia, Fernandez, Gonzalez), and this may create 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 the may do it with only their paternal or with their maternal surname, or they may hyphenate the two surnames. Authors may also sign using their first name, their middle name, or both.

We use the following matching procedure in order to deal with the above problems. First, we assign all publications and all professors in our sample to a broad disciplinary category. In order to attribute comparable disciplinary categories for publications and individuals, we aggregate disciplines defined by the Spanish Ministry and ISI disciplinary areas into the following categories: Agriculture; Chemistry; Biology; Geology; Physics; Mathematics and Computer Science; Engineering; Medicine, Veterinary and Pharmacology; Economics and Management; Psychology, Sociology and Political Science.[34] Second, in each broad disciplinary category we match publications with individuals in our database using the information on their surnames and initials.

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

---

[34]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.

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 the time elapsed between the publication date and the date when the number of received citations is observed. Therefore, we first normalize the number of citations received by each publication subtracting the average number of citations received by Spanish-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.

Teseo database on doctoral dissertations Since 1977 PhD candidates in Spanish universities register their dissertation in the database TESEO, which is run by the Ministry of Education. We retrieved all the information available in this database from the website `https://www.educacion.gob.es/teseo` in May 2011. While registration is compulsory, according to Fuentes and Arguimbau (2010) TESEO includes information on approximately 90% of all dissertations read in Spain during this period. We observe information on 151,483 dissertations. TESEO provides the identity and affiliation of dissertations' authors, advisors and committee members. Approximately 40% of dissertations are female authored. Female supervisors are scarce, only 18%. They supervise mostly female students: 58% of their students are women. The opposite is true for male advisors: 61% of their students are male.

We match TESEO data with the list of candidates and evaluators. In exams to full professor positions we are able to find the dissertation of 71% of candidates and 41% of evaluators. In exams to associate professor positions we observe the dissertation

of 83% of candidates and 70% of evaluators. Missing information may be due to the fact that individuals (i) did their PhD abroad, (ii) defended their dissertation before 1977, (iii) there are spelling mistakes, (iv) the dissertation was not included in TESEO for unknown reasons (approximately 10% of all dissertations), or (v) there was a homonymity problem (in our dataset 0.1% of individuals share the same name, middle name, paternal surname and maternal surname).

Each thesis has been classified by its author using the Unesco International Standard Nomenclature for Fields of Science and Technology. This is a system developed by Unesco that includes more than two thousand six-digits categories.[35] 80% of dissertations provide this information. Approximately half of the authors select one six-digit category, 35% select two categories, and 15% select three or more categories. There are on average around one hundred dissertations per category. We use this information to construct a measure of individuals' research interests. In particular, we take into account every dissertation where an individual appears as an advisor, committee member or author. We were able to obtain information on the research interests of 98% candidates to full professor positions, 94% of candidates to full professor positions, 98% of eligible full professors and 96% of eligible associate professors.

---

[35]Available at `http://unesdoc.unesco.org/images/0008/000829/082946eb.pdf`

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 were allowed to sit on a committee. In case that a second individual in one of these categories was drawn, the draw was rejected. 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 rules above. 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

49

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 positions:

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

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:[36]

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

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

For associate professor exams the probability of having zero, one, ..., or seven women in the committee is a bit more difficult to calculate. First, we compute the probability to draw zero, one, ..., or three women among full professors or senior researcher evaluators and zero, one, ..., or four women among associate professors 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 women among eligible full professors and associate professors, respectively, and M_1 and M_2 are the numbers of men 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:

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

---

[36]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.

Appendix D: Tables for the graphs

Table D1: Nonlinearities

	1	2
	Full professor exams	**Associate professor exams**
Female candidate	-0.026***	-0.004
	(0.009)	(0.009)
One female evaluator	-0.005	0.006
	(0.004)	(0.005)
Two female evaluators	-0.009	0.007
	(0.007)	(0.005)
Three female evaluators	-0.009	-0.001
	(0.009)	(0.008)
Four or more female evaluators	-0.035	0.016*
	(0.022)	(0.009)
Female candidate*One female evaluator	0.019	-0.016
	(0.013)	(0.013)
Female candidate*Two female evaluators	0.032	-0.024*
	(0.022)	(0.013)
Female candidate*Three female evaluators	0.034	-0.015
	(0.030)	(0.015)
Female candidate*Four or more female evaluators	0.083*	-0.046**
	(0.047)	(0.019)
Constant	0.114***	0.121***
	(0.003)	(0.004)
Candidates' characteristics	✓	✓
Adjusted R-squared	0.063	0.046
Number of observations	12680	17432

Notes: OLS estimates. Standard errors are clustered by exam. Candidates' characteristics include age, publications, citations per publication, number of students advised and the membership in PhD committees. All these controls are normalized at the exam level. All regressions include discipline fixed effects, probabilities to draw one, two, three, four, five, six or seven women, interactions of these probabilities with the female dummy, and the number of positions per candidate. These controls are rescaled to have zero mean in the corresponding subsample, therefore the constant is the average success rate of male candidates while the female dummy is the gender gap in the average committee.

Table D2: The effect of committees' gender composition on candidates' success, by travel time and feminization

	1	2	3	4
	Full professor exams		Associate professor exams	
	Travel time	Feminization	Travel time	Feminization
Male, t2	-0.034***	-0.000	-0.024***	0.016***
	(0.008)	(0.003)	(0.008)	(0.004)
Male, t3	-0.036***	0.003	-0.028***	0.004
	(0.008)	(0.004)	(0.010)	(0.006)
Female, t1	-0.029***	-0.026**	-0.015	-0.009
	(0.010)	(0.012)	(0.010)	(0.007)
Female, t2	-0.038***	-0.012*	-0.056***	-0.022***
	(0.010)	(0.007)	(0.009)	(0.005)
Female, t3	-0.061***	-0.012*	-0.056***	-0.015***
	(0.011)	(0.006)	(0.011)	(0.005)
Female FP evaluators	-0.035	-0.009	-0.106	-0.005
	(0.045)	(0.024)	(0.082)	(0.025)
Male, t2* Female FP evaluators	-0.038	-0.066*	0.173	0.001
	(0.072)	(0.038)	(0.115)	(0.056)
Male, t3* Female FP evaluators	0.023	-0.031	0.145	0.023
	(0.069)	(0.037)	(0.118)	(0.063)
Female, t1* Female FP evaluators	0.107	0.101	0.204	0.078
	(0.092)	(0.182)	(0.133)	(0.098)
Female, t2* Female FP evaluators	0.146*	0.223***	0.078	-0.027
	(0.083)	(0.078)	(0.112)	(0.066)
Female, t3* Female FP evaluators	0.109	0.068	0.019	-0.035
	(0.073)	(0.045)	(0.137)	(0.053)
Female AP evaluators			0.110**	0.013
			(0.048)	(0.017)
Male, t2* Female AP evaluators			-0.158**	0.019
			(0.066)	(0.031)
Male, t3* Female AP evaluators			-0.066	0.013
			(0.074)	(0.048)
Female, t1* Female AP evaluators			-0.135*	-0.104
			(0.080)	(0.063)
Female, t2* Female AP evaluators			-0.136**	-0.040
			(0.065)	(0.035)
Female, t3* Female AP evaluators			-0.166**	-0.047
			(0.083)	(0.036)
Constant	0.134***	0.110***	0.143***	0.120***
	(0.005)	(0.002)	(0.006)	(0.002)
Adjusted R-squared	0.024	0.032	0.030	0.033
Number of observations	13601	13601	14890	18139

Notes: OLS estimates. Standard errors are clustered by exam. Terciles of travel time are defined according to the car travel time from the institution where a candidate is based to the location of the exam, in hours: [0, 3.55], (3.55, 6.33], (6.33, 12]. Terciles of feminization are defined according to the proportion of women among tenured professors in the discipline: [0, 0.21], (0.21, 0.34], (0.34, 0.65]. Columns (1) and (3) include discipline fixed effects and all regressions include the number of positions per candidate. These controls are rescaled in such a way that the constant is the average success rate of male candidates in the first tercile. FP and AP stand for full and associate professor respectively. Female FP/AP evaluator is the difference between the actual and the expected proportion of female evaluators (of a given rank) among all committee members.