

# Does the Gender Composition of Scientific Committees Matter?

By MANUEL BAGUES, MAURO SYLOS-LABINI, AND NATALIA ZINOVYEVA\*

*We analyze how a larger presence of female evaluators affects committee decision-making using information on 100,000 applications to associate and full professorships in Italy and Spain. These applications were assessed by 8,000 randomly selected evaluators. A larger number of women in evaluation committees does not increase either the quantity or the quality of female candidates who qualify. Information from individual voting reports suggests that female evaluators are not significantly more favorable towards female candidates. At the same time, male evaluators become less favorable towards female candidates as soon as a female evaluator joins the committee.*

*JEL: J71, J16*

*Keywords: Scientific committees, gender discrimination, randomized natural experiment.*

The underrepresentation of women in academia remains a cause for concern among universities and policy makers around the world. In Europe, women account for 47 percent of PhD graduates, 37 percent of associate professors and only a mere 21 percent of full professors (European Commission, 2016). Similar patterns may be observed in the US and the gender imbalance is even larger in Japan (National Research Council, 2009, Abe, 2012).

Several explanations may account for the lack of women in high-level positions. According to the pipeline theory, once women have entered the lower rungs of the academic career it is mainly a matter of time that they would move their way through a metaphorical pipeline to reach high-level jobs. However, in most disciplines, the share of women among faculty members remains low even after decades of improved recruitment of women at the undergraduate and the doctoral level (Ginther and Kahn, 2004, 2009). Gender differences in promotion

\* Bagues: Aalto University, Department of Economics, P.O. Box 21240, FI-00076 AALTO, Helsinki, Finland, manuel.bagues@aalto.fi. Sylos-Labini: University of Pisa, Department of Political Science, Via Serafini 3, Pisa, Italy, mauro.syloslabini@unipi.it. Zinovyeva: Aalto University, Department of Economics, P.O. Box 21240, FI-00076 AALTO, Helsinki, Finland, natalia.zinovyeva@aalto.fi. This paper combines information that was previously reported in two separate working papers: Bagues, Sylos-Labini and Zinovyeva (2014) and Zinovyeva and Bagues (2011). We would like to thank Olympia Bover, Irma Clots-Figueras, Sara de la Rica, Gemma Derrick, Juan Jose Dolado, David Dorn, Silvia Dorn, Berta Esteve-Volart, Luis Garicano, Marco Giarratana, Elena Martínez, Nic Morgan, Javier Ruiz-Castillo and participants in numerous presentations for their useful comments. We also acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada, Instituto de la Mujer (research grant 36/12) and the Spanish Ministry of Science and Technology (research grants ECO2008-06395-C05-05 and ECO2008-01116). The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

rates might also reflect differences in productivity, perhaps due to the existence of gendered roles at the household level or the lack of female mentors and role models (Blau et al., 2010). Some women may also devote excessive time to tasks that are socially desirable but which are not taken into account in promotion decisions (Vesterlund, Babcock and Weingart, 2014). Furthermore, some authors have pointed out that women are less likely to apply for promotions (Bosquet, Combes and Garcia-Peñalosa, 2013, De Paola, Ponzio and Scoppa, 2015), perhaps due to the existence of gender differences in the preference for competitive environments (Niederle and Vesterlund, 2007, Buser, Niederle and Oosterbeek, 2014) or in bargaining abilities in the labour market (Babcock et al., 2006, Blackaby, Booth and Frank, 2005).

Beyond these supply-side explanations, the slow progress made by women has been sometimes attributed to the lack of female evaluators in the committees which decide on hiring and promotions. In this paper we examine whether the presence of women in scientific committees might help to increase the chances of success of female candidates and to improve the quality of the evaluations. There are several reasons for considering this hypothesis. First, there is evidence of gender segregation across different scientific subfields (Dolado, Felgueroso and Almunia, 2012, Hale and Regev, 2014). If men and women tend to do research in different subfields and evaluators overrate the importance of their own types of research, the lack of female evaluators might be detrimental for female candidates (Bagues and Perez-Villadoniga, 2012, 2013). Second, research networks tend to be gendered (Boschini and Sjögren, 2007, Hilmer and Hilmer, 2007).<sup>1</sup> If evaluators are mostly male, male candidates might have a better chance to be acquainted with committee members and could perhaps benefit from these connections (Zinovyeva and Bagues, 2015, Bagues, Sylos-Labini and Zinovyeva, 2015). Third, men might hold more negative stereotypes of women than other women do or they may be biased against women reaching high-level positions. For instance, according to the World Value Survey, around 25 percent of US males believe that men make better political leaders and 16 percent think that men make better business executives. Women are half as likely to hold such views. A similar pattern is observed in Europe.<sup>2</sup> According to some authors, similar biases are also present in the academic world.<sup>3</sup> Fourth, the presence of women in

<sup>1</sup>Boschini and Sjögren (2007) show that coauthoring is not gender neutral in Economics. Hilmer and Hilmer (2007) observe that in the US 55 percent of the Economics PhD students being advised by women are female, while only 18 percent of Economics PhD students advised by men are female.

<sup>2</sup>World Value Survey Wave 6: 2010-2014. Official aggregate v.20140429. World Values Survey Association ([www.worldvaluessurvey.org](http://www.worldvaluessurvey.org)).

<sup>3</sup>Gender discrimination in academia remains a controversial issue. According to meta-analyses by Ceci and Williams (2011) and Ceci et al. (2014), the more recent empirical evidence fails to provide any clear support to the assertions of discrimination in manuscript reviewing, interviewing, and hiring. However, other studies find that female researchers might still receive lower evaluations than male researchers with identical characteristics (Steinpreis, Anders and Ritzke, 1999, Moss-Racusin et al., 2012). Some experts in gender studies have also argued that male evaluators discriminate against female candidates. For instance, in a report commissioned by the European Commission, the expert group *Women In Research Decision Making* concludes that "(a)t the very least, having male only committees risks replicating

evaluation committees might also improve the quality of the evaluation. It has been argued that group performance is positively correlated with the proportion of women in a group (Woolley et al., 2010). The presence of women in scientific boards might not only help to achieve gender balance in the academic profession, but it can also make science more meritocratic and invigorate its progress.

These arguments seem to have reached policy makers. A number of countries have introduced quotas requiring that scientific committees be at least 40 percent female (and male) and many universities and scientific institutions have their own internal guidelines ensuring the presence of both genders in committees.<sup>4</sup> However, despite the increasing popularity of gender quotas in scientific committees, there are concerns about their effectiveness. Quotas are costly for senior female researchers, as they increase disproportionately the amount of time that they have to devote to evaluation committees (Vernos, 2013). Furthermore, a larger presence of women in committees may not necessarily benefit female candidates. Both men and women have developed their careers in an academic environment dominated by men, and both genders may tend to associate important academic positions, and the features they require, with men, not with women (Méndez and Busenbark, 2015). And even if women are relatively more sympathetic towards female candidates, they may not have equal levels of voice and authority in deliberation processes (Karpowitz, Mendelberg and Shaker, 2012, Brescoll, 2011). The presence of women in the committee can also induce male evaluators to be less favorable towards female candidates. Past research on group dynamics suggests that men might not respond favorably to the presence of gender diversity, particularly in domains that men have historically dominated (Crocker and McGraw, 1984). Female evaluators can also contribute to strengthen male identity (Akerlof and Kranton, 2000) or they can trigger a licensing effect (Monin and Miller, 2001, Khan and Dhar, 2006).

A better understanding of the impact of scientific committees' gender composition on recruitment and promotion decisions is crucial in order to determine whether quotas are desirable. The empirical evidence has been so far inconclusive and typically based on small samples. Sometimes researchers seem to benefit from the presence of evaluators who share the same gender (Casadevall and Handelsman, 2014, De Paola and Scoppa, 2015), sometimes they seem to obtain relatively

stereotypes and bias, both regarding applicants and issues in research" (European Commission, 2008, page 27). Another expert report on the situation of women researchers in Spain asserts that "there are prejudices about women among those who co-opt, promote or have the key to promotion. The bodies which control this are mostly male and, even if they are not totally conscious of it, they see an academic woman first as a woman and secondly as a colleague." (Fundación Española para la Ciencia y la Tecnología, 2005, page 48). Other researchers have voiced similar views (Bagilhole, 2005, Barres, 2006, Smith et al., 2015).

<sup>4</sup>Gender quotas in scientific committees were introduced in 1995 in Finland (amendment of the *Act on Equality between Women and Men*, Act No. 624/1992 and No. 206/1995.), in 2007 in Spain (*Constitutional Act 3/2007 of 22 March for Effective Equality between Women and Men*) and in 2014 in France (decree No. 2014-997, September 2 2014). The European Commission has also committed to reaching a target of 40 percent female participation in its advisory structures for Horizon 2020, the European Union's research and innovation programme for 2014-20.

better evaluations from opposite-sex evaluators (Broder, 1993, Ellemers et al., 2004), and in some other cases gender does not seem to play any (statistically) significant role (Abrevaya and Hamermesh, 2012, Jayasinghe, Marsh and Bond, 2003, Milkman, Akinola and Chugh, 2015, Moss-Racusin et al., 2012, Steinpreis, Anders and Ritzke, 1999, Williams and Ceci, 2015). A brief summary of these studies is available in Appendix A.<sup>5</sup> It is unclear whether these mixed findings reflect the idiosyncrasies of the different situations and samples analyzed in each study, or simple random sampling variation. The literature also does not shed light on the mechanisms through which a higher presence of female evaluators in committees may benefit female candidates. From a policy perspective, the lack of more extensive and clear evidence is disappointing.<sup>6</sup>

In this paper we analyze the role of evaluators' gender in academic evaluations using the exceptional evidence provided by two large-scale randomized natural experiments in two different countries, Italy and Spain. The representation of women in Italian and Spanish universities is similar to their representation in other European countries and the US. Despite having achieved parity at the lower rungs of the academic ladder, women are still underrepresented in top academic positions. They account for approximately half of new PhD graduates, one third of associate professors but only one fifth of full professors in both countries.<sup>7</sup> The Spanish and Italian institutional arrangements offer several unique features. In order to be either promoted or hired by a university at the level of associate or full professor, researchers are required to first obtain a qualification granted by a centralized committee at the national level. In these nation-wide examinations, which are performed periodically in all disciplines in both countries, evaluators are selected from a pool of eligible professors using a random draw. This allows us to consistently estimate the causal effect of committees' gender composition on evaluations. We also observe extensive and detailed information on evaluators' and candidates' research production, academic connections and their subfield of specialization. We exploit this information to explore the different mechanisms suggested by the theory about the role of committees' gender composition. Each country also offers some comparative advantage in terms of data availability. We use individual voting reports, available in Italy, to study the voting behavior of male and female evaluators within each committee. In Spain, we can observe the

<sup>5</sup>All appendix material can be found in the Online Appendix.

<sup>6</sup>A related literature also analyses the role of evaluators' gender in nonacademic occupations (Bagues and Esteve-Volart, 2010, Bertrand et al., 2014, Booth and Leigh, 2010, Kunze and Miller, 2014), in sport activities (Sandberg, 2014) or in the lab (Bohnet, van Geen and Bazerman, 2015). In general, in these studies evaluators' gender is not relevant, with the exception of Bagues and Esteve-Volart (2010) who document that female applicants to the Spanish judiciary have lower chances of being hired when they are randomly assigned to an evaluation committee including women.

<sup>7</sup>In Italy, women account for 52 percent of new PhD graduates, 36 percent of associate professors, and 21 percent of full professors (MIUR - Italian Ministry of Education, University and Research, 2016). In Spain, women account for 49 percent of new PhD graduates, 40 percent of associate professors, and 21 percent of full professors (MECD - Spanish Ministry of Education, Culture and Sports, 2016). According to information from individuals who obtained a PhD in the 90s in Spain, female graduates are half as likely to attain full professorship than male graduates (Sánchez de Madariaga, de la Rica and Dolado, 2011).

future productivity of promoted candidates. We use this information to examine the quality of the assessments granted by committees with different gender compositions. As we explain in more detail in section I, there exist also a number of interesting institutional differences between the evaluation processes in the two countries. Having data for the two different institutional arrangements allows us to cross-validate the findings and to explore their robustness.

Our database includes information on all qualification exams that were conducted in Italy in years 2012-2014 and in Spain in years 2002-2006. Overall, these evaluations involved approximately 100,000 applications and 8,000 evaluators in all disciplines. Evaluation committees, which include five members in Italy and seven members in Spain, are composed mostly by men. Approximately one third of evaluation committees do not include any women, in one third there is just one female evaluator, and in one third of committees there are two or more women, but very rarely we observe a female majority.

In both countries male applicants tend to be more successful than female applicants. In Italy, approximately 38 percent of men receive a positive evaluation, compared to 35 percent of women. In Spain, 12 percent of male applicants qualify, while the success rate among female applicants is equal to 11 percent. When we take into account candidates' observable productivity, the remaining gender gap is equal to 1.5 percentage points (p.p.) in Italy and 1.4 p.p. in Spain, and it is statistically significant in both countries. We find no empirical support, neither from the average in the two countries nor from the majority of subsamples analyzed, to suggest that the presence of women in evaluation committees decreases the gender gap in a statistically or economically significant way. On the contrary, in Italy gender-mixed committees exhibit a significantly larger gender gap than committees composed only of male evaluators. An extra woman in a committee of five members increases the gender gap by somewhere between 0.4 and 3.3 p.p., considering a 95 percent confidence interval. In the Spanish case, we can reject any sizable impact. An additional woman in a committee of seven members may decrease the gender gap by at most 0.5 p.p. or it might also increase it by up to 1 p.p.

We also examine whether committees with a relatively larger proportion of women promote better candidates, using as a proxy of candidates' quality their research output before the evaluation and, in the case of Spain, also their research output during the following five years. We do not observe any significant difference in the past or future observable quality of candidates who have qualified in committees with different gender compositions.

Evidence from 300,000 individual voting reports, available in the case of Italy, suggests that there are two main factors that explain why female candidates do not benefit from a larger presence of women in committees. In mixed gender committees, female evaluators rate female applicants higher than their male colleagues, but the difference is small and statistically non significant. At the same time, the presence of female evaluators in committees makes male evaluators tougher upon

female candidates, perhaps reflecting a licensing effect or male identity priming.

To gain a better understanding of why female evaluators do not exhibit a stronger same-sex preference and also to determine the validity of our findings in other contexts, we explore why none of the standard theories predicting that a larger presence of women in committees helps female candidates plays a major role in this context. First, we consider the gendered networks hypothesis. As expected, we find that research networks tend to be gendered in both countries. Female professors are significantly more likely to have an advisor, a colleague or a coauthor of the same gender. We also observe that committees tend to favor connected candidates. However, the likelihood of having a connection in a national committee is relatively low and, therefore, networks have only a limited effect on the evaluation outcomes. Second, we examine the role of gender segregation across research subfields. At the level at which evaluations were conducted, around 200 different fields, gender segregation turns out to be relatively small. As a result, while evaluators tend to prefer candidates with a similar research profile, the impact of gender segregation on evaluations is negligible. Third, we study gender stereotypes. Stereotypes are expected to be more relevant when evaluators cannot observe accurately the quality of candidates, for instance because evaluators and candidates are specialized in different subfields of research. The influence of stereotypes on evaluation outcomes seems to increase, not decrease, when there are women in the committee. Finally, we also examine separately evaluations for high-level positions. Male evaluators might have prejudices against women being promoted to full professorships, but not to positions at lower levels of the career ladder. Results are mixed: we find support for this hypothesis in the case of Spain, but not in the case of Italy.

Our study contributes to the literature in several ways. We provide the first large-scale assessment of the causal impact of the gender composition of scientific committees. There is no evidence suggesting that, in the two evaluation systems considered in this study, female candidates benefit from the presence of a larger share of women in evaluation committees. We also examine explicitly the relevance of the different theoretical arguments that have been proposed in the literature in favor of increasing the share of women in committees. This analysis helps to assess the external validity of our findings and, as we discuss in detail in the final section of the paper, it provides a better understanding of when gender quotas might be desirable. Finally, we open the black box of committee decision-making and we analyze the voting behavior of individual committee members. Our findings suggest that interactions within committees might exacerbate the impact of gender stereotypes.

## I. Institutional background

Several European countries have national evaluation systems which are meant to guarantee the academic quality of professors in public universities. The evidence presented in this paper is based on an analysis of two variants of such

systems: the Italian system known as *Abilitazione Scientifica Nazionale*, which was introduced in 2012, and the Spanish system known as *Habilitación*, which was in place between 2002 and 2006.

Both systems require candidates for associate and full professorships to qualify in national evaluations held by an academic board in the appropriate discipline. In each country, there are nearly two hundred legally defined academic disciplines, each corresponding to a certain area of knowledge. Successful candidates can then apply for a position at a given university. The time line of evaluations has the following steps. First, a call for applicants is announced in which candidates can apply for multiple disciplines and positions. When the list of applicants is settled, committee members are randomly selected from the list of eligible evaluators in the corresponding discipline. Once the committees are formed, the evaluation process begins and once this is over, the evaluation results are made public. Rostered evaluators can potentially resign at any point of the process, something that happens in 2 percent of cases in Spain and in 8 percent of cases in Italy. Resigned evaluators are substituted by randomly selected evaluators.

The procedure has also distinctive features specific to each country. In Spain, evaluations involve oral presentations by the candidates, while in Italy evaluations are based only on candidates' CVs and publications. In Spain qualification leads almost automatically to promotion, while in Italy the chances to get promoted conditional on obtaining qualification are much lower. The Italian system is relatively more transparent and exposed to public scrutiny. Nonetheless, in both systems there seems to be room for subjectivity. For instance, Zinovyeva and Bagues (2015) and Bagues, Sylos-Labini and Zinovyeva (2015) document that the presence of a coauthor or a colleague in the evaluation committee has a significant positive impact on candidates' chances of success in both countries.

We describe in detail the main features of each system below. This information is also summarized in Appendix B.

#### A. *Abilitazione Scientifica Nazionale*

In Italy, four out of five committee members are selected through a random draw from the pool of 'Italian' eligible evaluators and the remaining evaluator is drawn out of the pool of 'foreign' eligible evaluators. The former pool consists of full professors affiliated to Italian universities who volunteered to be members. The latter pool consists of professors affiliated to universities from OECD countries, who also voluntarily participate in Italian evaluations. The randomization procedure is subject to one important constraint: no university can have more than one evaluator within a single committee.

The eligibility of evaluators is decided in the following way. In science, technology, engineering, mathematics, medicine and psychology (STEMM), evaluators are required to have a research output above the median for full professors in the discipline in at least two of the following three dimensions: (i) the number of articles published in scientific journals, (ii) the number of citations, (iii) and the H-

index. In the social sciences and the humanities (SSH), the research performance of evaluators has to be above the median in at least one of the following three dimensions: (i) the number of articles published in high quality scientific journals (in what follows, A-journals),<sup>8</sup> (ii) the overall number of articles published in any scientific journals and book chapters, and (iii) the number of published books. ‘Foreign’ eligible evaluators have to satisfy the same requirements. While ‘Italian’ evaluators work *pro bono*, ‘OECD’ evaluators receive €16,000 for their participation.

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

An important feature of the Italian system is its extreme transparency: all the relevant information – including candidates’ and evaluators’ CVs, as well as individual evaluation reports – is published online. An independent evaluation agency appointed by the ministry also collects and publicizes information on the research output of final candidates in the ten years preceding the evaluation, as measured by the three bibliometric indicators described above. The evaluation agency compared the research productivity of candidates in each of these three dimensions with the research productivity of professors in the category to which they applied, and committees were asked to take this information into consideration.

### B. *Habilitación*

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

In order to be eligible, evaluators are required to satisfy some minimum research level which is assessed by the Spanish education authority.<sup>9</sup> This requirement is

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

<sup>9</sup>The Spanish education authority determines professors’ eligibility according to the number of *sexenios* completed. *Sexenios* are granted periodically by the ministry on the basis of applicants’ research



satisfied by approximately 81 percent of full professors and 70 percent of associate professors. Unlike the Italian system, where participation is voluntary, in Spain all eligible professors can be selected to serve in committees.

Candidates for evaluation are required to make several oral presentations in front of a committee. For candidates to full professorships, these exams have two qualifying stages. In the first stage, each candidate presents the CV and then, in the second stage, an example of his or her research work. Exams for the position of associate professor, in addition to these two stages, have an intermediate stage where candidates give a lecture on a topic randomly chosen from a syllabus proposed by the candidate. In each stage evaluations are made on a majority basis. Qualifications have unlimited validity once they have been granted. The number of qualifications conceded at the national level is very limited and being accredited is, in most cases, equivalent to being promoted.

## II. Data

We use data on all evaluations from the first edition of the Italian *Abilitazione Scientifica Nazionale* (years 2012-2014) and on all evaluations from the Spanish *Habilitación* (years 2002-2006). In Italy, the data includes information on 184 committees, one per each academic discipline. Each committee assessed both applications to associate and to full professorships. In Spain, there are in total 967 committees in 174 disciplines, of which 502 are committees evaluating candidates for full professorships and 465 evaluating candidates for associate professorships.

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

### A. Evaluators

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

output in any uninterrupted period of a maximum of six years. Eligible associate professors are required to have held at least one *sexenio* while eligible full professors are required to have held at least two *sexenios*.

expected share of women in the committee is around 18 percent, which is similar to the initial share of women in actual committees.<sup>10</sup> Approximately one out of every thirteen evaluators resigned and was replaced by another eligible evaluator. These replacements slightly increased the share of women in committees to 19 percent, but the difference is not statistically significant. 41 percent of committees include no women at all, in 35 percent there is one woman, in 16 percent there are two women, and only 8 percent of committees have a majority of female evaluators.

Table C1 in Appendix C provides descriptive information on eligible evaluators based in Italy.<sup>11</sup> On average, they have been in a full professor position for 13 years. They list 131 publications in their CVs, of which just over half are articles in scientific journals, and the rest are books, book chapters, publications in conference proceedings, patents, etc. Around half of these publications were published during the previous ten years. To assess the quality of research output, in STEMM disciplines we compute their total Article Influence Score, summing up the Article Influence Score of all publications; in SSH disciplines we use the number of articles in A-journals.<sup>12</sup> About 28 percent of eligible professors are based in the South of Italy.

In columns 2-4, we compare characteristics of male and female evaluators. For this comparison, we normalize all variables at the discipline level. Female evaluators have significantly shorter tenure than their male counterparts and they also have lower research output in almost all dimensions. They are less likely to be based in the South, but this difference is not significant.

In Spain, the lists of eligible evaluators include 49,199 full professors and 61,052 associate professors.<sup>13</sup> Women constitute 35 percent of eligible associate professors, but only 14 percent of full professors are women. Taking into account the composition of both pools, the expected share of women in the committee is around 19 percent. This figure is similar to the share of women in the initial set of committees selected by random draw and is unaffected by the resignation of 2 percent of evaluators. Overall, 32 percent of committees are composed by only male evaluators, 29 percent of committees have one woman on board, 22 percent include two women, 11 percent three women, while only 6 percent have more women than men.

We collect information on the research outcomes of Spanish researchers from several sources. We observe their publications in international journals covered by Web of Knowledge and their articles and books in the Spanish language included

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

<sup>11</sup>Unfortunately, we were unable to gather systematic information on 'foreign' evaluators. In their case, the official CVs are not in a standardized format and they are often incomplete.

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

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

in the database Dialnet, as well as patents in the European Patent Office in which these researchers are listed as inventors. We also have information on their activity as Ph.D. advisors and as members of dissertation committees. We compare female and male eligible evaluators, normalizing their characteristics at the level of exam and category. Results are very similar to the ones observed for the Italian academia (see columns 6-8 and 10-12 of Table C1). Female eligible evaluators are younger, have shorter tenure, and on average they published less than male researchers in the same discipline and rank. They have also lower accumulated quality-adjusted scientific production, they tend to participate less in advising and evaluating doctoral students, and they are relatively less likely to come from universities located in the southern regions of the country.<sup>14</sup>

### B. Candidates

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

As shown in the upper panel of Table C2, 31 percent of applications for the position of full professor and 41 percent of applications for the position of associate professor were submitted by women. Candidates for a full professorship are about 49 years old and candidates for an associate professorship are six years younger. About half of the applicants for associate professorships hold a permanent contract and about three fourths of applicants for full professorships do. Candidates mainly apply for an evaluation in the field in which they currently hold a permanent contract.

Female applicants tend to be younger among applicants for associate professorships, and they are of a similar age as their male counterparts in evaluations for full professorships (columns 3-5 and 8-10). In both cases the publication record of female candidates is significantly weaker. The only dimension in which women seem to be achieving better results than men is in publishing conference proceedings. In addition to information on productivity coming from candidates' CVs, we observe the order in which candidates submitted their applications. In principle, the timing of the application might reflect both candidates' self-confidence and quality. We normalize this variable uniformly between 0 and 1. We observe that female candidates for the post of full professor apply a bit later than their male counterparts, but no similar gender difference can be observed among candidates for associate professor positions. In Italy, approximately 14 percent of applications were withdrawn once the identity and the criteria of evaluators were made public. Withdrawals were more common among female applicants. Overall,

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

approximately 38 percent of applications by male candidates and 35 percent of applications by female candidates were successful.

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

In addition to the final decision of the committee, we also collected information on the individual evaluation reports, available in the case of Italy. Overall we observe around 300,000 individual reports. 45 percent of these reports were favorable to the candidate and most of the time decisions were taken unanimously (in 86 percent of the cases). Unanimity is relatively more frequent when applicants are below the median in each of the three corresponding bibliometric dimensions (93 percent) and when applicants are above the median in all three dimensions (86 percent), and it is lower when applicants are above the median only in one (84 percent) or two dimensions (82 percent).

In Spain, overall there were 13,444 applications for full professorships and 17,799 applications for associate professorships (lower panel of Table C2). The gender ratios among applicants are very similar to the ones in Italy: around 27 percent of applicants to full professor are women and there are around 40 percent of women among applicants to associate professor. Once again, male applicants seem to have stronger research records than their female counterparts. They also tend to be slightly more successful in evaluations.

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

### *C. Connections*

We identify professional links between candidates and eligible evaluators. We consider all the possible interactions within each discipline, around 2.5 million possible pairs in Italy and 5.5 millions in Spain. As shown in Table C3, the probability that a candidate and an eligible evaluator are affiliated to the same institution is around 3 percent in Italy and 5 percent in Spain. The probability that they have coauthored a paper is smaller: 1.4 percent in Italy and 0.4 percent in Spain.

In the case of Spain, we also observe if there was a student-advisor relationship

or if the candidate and the eligible evaluator have participated in the same thesis committee.<sup>15</sup> These links are relatively rare: in 0.2 percent of the cases the eligible evaluator is the PhD thesis director of the candidate and in 1.3 percent they have participated in the same thesis committee.

Male candidates tend to have more coauthors among eligible evaluators and they are more likely to have interacted with an eligible evaluator previously in a thesis committee (Table C3, columns 3-5).

#### D. Research similarity

We also collect information on the overlap of research interests between candidates and eligible evaluators. Due to data availability, there are some differences in how we define research similarity in the two countries. In the case of Italy, we have information on the field and the subfield where researchers with a permanent contract in an Italian university are officially registered. There are 184 fields (*settore concorsuale*) and approximately 370 subfields (*settore scientifico-disciplinare*).<sup>16</sup> In about 60 percent of the cases the candidate and the eligible evaluator belong to the same subfield (Table C3).

In the case of Spanish researchers, we infer their research interests using information on their participation in doctoral dissertations, either as authors, advisors, or committee members. In Spain, all doctoral theses are classified in more than two thousand categories.<sup>17</sup> Economics, for example, is divided into one hundred different research fields (e.g.: Labor Economics). We construct a measure of the overlap of the research interests of candidates and evaluators based on the subfield of every dissertation where they have been involved. In the spirit of Jaffe (1986) and Bloom, Schankerman and Van Reenen (2013), we measure research proximity between individuals  $i$  and  $j$  as the angular separation of the vectors  $S_i = (S_{1i} \dots S_{Ci})$  and  $S_j = (S_{1j} \dots S_{Cj})$ , where  $S_{Ci}$  is the share of dissertations in category  $C$  in which individual  $i$  has been involved:

$$(1) \quad \text{Overlap}_{ij} = \frac{S_i S_j'}{(S_i S_i')^{1/2} (S_j S_j')^{1/2}}.$$

This index takes value one if two individuals have participated in dissertations in the same subfields in the same proportion and value zero if there is no overlap. On average, in our sample the degree of overlap between candidates and evaluators

<sup>15</sup>We consider three possible interactions: (i) the evaluator was a member of candidate's thesis committee, (ii) one of them had invited the other to sit in her students' thesis committee, or (iii) both of them sat in the same student thesis committee.

<sup>16</sup>Historically, each Italian researcher was assigned to certain *settore scientifico-disciplinare*. More recently, upon the introduction of the new system of competitive exams, researchers were assigned also to a *settore concorsuale*. The correspondence between the two classifications is not always unique, in some cases researchers belonging to the same *settore scientifico-disciplinare* may be assigned to different *settore concorsuale*.

<sup>17</sup>The author of the dissertation selects the subfield using the *International Standard Nomenclature for Fields of Science and Technology*, a system developed by Unesco.

is equal to 0.20. As shown in Table C3, female candidates are slightly more likely than male candidates to share their research interests with eligible evaluators.

### III. Empirical analysis

We start our analysis by providing descriptive information on the average success rate of male and female applicants, unconditional and conditional on their observable research productivity. Then we investigate how the gender composition of committees affects the success rate of male and female candidates, candidates' decision to withdraw their application and the quality of male and female applicants who qualify. To achieve a better understanding of the observed patterns, we use the information provided by individual voting reports to examine how male and female evaluators vote within the same committee. We examine whether male and female evaluators vote differently depending on the gender of applicants, and we also investigate whether the presence of women in a committee affects the voting behavior of male evaluators. Finally, we explore the relevance of the main theories according to which evaluators' gender may be relevant.

#### A. Gender gap

We estimate the gender gap separately for the applicants in the two countries using the ordinary least squares (OLS) method:

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

where  $Y_{ie}$  is a dummy variable that takes value one if candidate  $i$  qualifies in evaluation  $e$  and takes value zero if the candidate receives a negative evaluation or withdraws the application before receiving the evaluation. Each evaluation  $e$  refers to the examination that was conducted in a given field and position (e.g. qualification for an associate professorship in *Applied Economics* in Spain in year 2005).  $Female_i$  is a dummy variable indicating the gender of the candidate and  $\mathbf{X}_i$  includes all (normalized) productivity indicators and individual characteristics listed in Table C2. We allow the effect of productivity indicators to vary across disciplinary groups, and the effect of age and contract type to vary across disciplinary groups and levels of promotion. Evaluation fixed effects ( $\mu_e$ ) control for any differences across evaluations that might affect the success rate of male and female candidates in a similar way. Throughout the analysis, we cluster standard errors at the committee level.

In Italy, the success rate of female candidates is 2.8 p.p. lower than male candidates in the same exam, unconditional on any measure of quality (Table 1, column 1, upper panel). In Spain, the unconditional gender gap is equal to 2.2 p.p. (column 1, lower panel). In both countries, approximately half of the gender gap can be explained by the differences in observable characteristics (column 2). The remaining conditional gender gap is equal to 1.5 p.p. in Italy (4 percent

relative to the success rate of men) and 1.4 p.p. in Spain (12 percent) and it is statistically significant in both countries.

It is unclear whether the remaining gap should be attributed to evaluators biases or to differences in unobservable characteristics. There may be substantial differences in the quality of male and female candidates which are not fully captured by our controls. Furthermore, the individual proxies of quality that we use in our analysis, such as position, affiliation or publications might also be the outcome of discriminatory processes, which would further hinder the interpretation of  $\beta_1$ .

*B. The impact of committees' gender composition on the chances of success of male and female candidates*

We examine whether the gender composition of committees affects the success rate of male and female applicants. In order to obtain causal estimates, our analysis exploits the random assignment of evaluators to committees. We compare the performance of applicants who initially were expected to face an evaluation committee with the same gender composition but, due to the random draw, were assigned to committees with a different number of female evaluators. Given that a few of the evaluators who were initially selected eventually declined to participate and were substituted by other (randomly selected) evaluators, first we report results from an intention-to-treat analysis where our independent variable is the gender composition of the initial set of evaluators. Later on, we instrument the gender composition of the committee which actually evaluated applicants using the gender composition of the committee initially drawn.

INTENTION-TO-TREAT ANALYSIS

We estimate the following equation on the pool of applicants using OLS:<sup>18</sup>

$$(3) \quad Y_{ie} = \beta_0 + \beta_1 Female_i + \beta_2 Female_e^{initial} + \beta_3 Female_i * Female_e^{initial} + \beta_4 Female_e^{expected} + \beta_5 Female_i * Female_e^{expected} + \mathbf{X}_i \beta_6 + \epsilon_{ie}$$

where  $Female_e^{initial}$  represents the share of female evaluators in the committee that was initially randomly drawn, before any evaluator resigned, and  $Female_e^{expected}$  is the expected share of women in this committee, calculated based on the composition of the pool of eligible evaluators and the rules that determine the draw.<sup>19</sup> In order to increase the accuracy of the estimation, we also include applicants' predetermined characteristics ( $\mathbf{X}_i$ ) and, in some specifications, evaluation fixed effects ( $\mu_e$ ).

<sup>18</sup>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.

<sup>19</sup>To ease the interpretation of coefficient  $\beta_1$ , we center  $Female_e^{expected}$  at zero by subtracting its sample mean.

TABLE 1—THE CAUSAL IMPACT OF COMMITTEES' GENDER COMPOSITION

<i>Dependent variable:</i>	1	2	3	4	5	6	7
	Qualified			Applied			
	OLS	OLS	ITT	IV	IV	IV	IV
	Italy						
Female candidate	-0.028 (0.006)	-0.015 (0.005)	-0.004 (0.009)	0.001 (0.011)	0.008 (0.008)	0.009 (0.007)	-0.026 (0.006)
Share of women in committee			0.000 (0.059)	-0.0004 (0.071)	-	-	-
Female candidate* Share of women in committee			-0.092 (0.036)	-0.116 (0.050)	-0.128 (0.035)	-0.132 (0.036)	-0.025 (0.026)
Adj. R-squared	0.001	0.240	0.245	0.245	0.236	0.236	0.075
Number of observations	69020	69020	69020	69020	69020	69020	69020
Mean dep. var. (for men)	0.38	0.38	0.38	0.38	0.38	0.38	0.87
	Spain						
Female candidate	-0.022 (0.004)	-0.014 (0.004)	-0.009 (0.007)	-0.009 (0.007)	-0.011 (0.007)	-0.010 (0.007)	
Share of women in committee			0.011 (0.017)	0.012 (0.018)	-	-	
Female candidate* Share of women in committee			-0.018 (0.026)	-0.019 (0.027)	-0.016 (0.028)	-0.022 (0.028)	
Adj. R-squared	0.001	0.036	0.039	0.039	0.005	0.005	
Number of observations	31243	31243	31243	31243	31243	31243	
Mean dep. var. (for men)	0.12	0.12	0.12	0.12	0.12	0.12	
<i>Controls for both panels:</i>							
Candidate characteristics		Yes	Yes	Yes	Yes	Yes	Yes
Exam FE	Yes	Yes			Yes	Yes	Yes
Expected share of women			Yes	Yes	Yes	Yes	Yes
Female candidate* Expected share of women			Yes	Yes	Yes	Yes	Yes
Committee characteristics						Yes	

*Note:* *Candidate characteristics* include all individual predetermined characteristics listed in Table C2 in the Online Appendix. *Committee characteristics* include the interaction between candidates' gender and the average tenure of evaluators (Italy only), their age (Spain only), their quality-adjusted productivity during the previous 10 years, and the proportion of committee members based in the South. The first-stage results for the IV estimations reported in columns 4 and 5 are available in Table D1. Standard errors are clustered by committee.

Coefficient  $\beta_2$  captures the causal effect of committees' initial gender composition upon the success rate of male candidates, and coefficient  $\beta_3$  shows how the gender gap varies depending on the share of women in the committee. Since  $Female_e^{initial}$  is computed using the initial assignment of evaluators, coefficients  $\beta_2$  and  $\beta_3$  provide intention-to-treat estimates. The causal interpretation of  $\beta_2$  and  $\beta_3$  relies on the assumption that the assignment was indeed random. The way in which the randomization was conducted in each country suggests that there was



little room for manipulation.<sup>20</sup> Nonetheless, before moving into the discussion of the impact of committees' gender composition on candidates' chances of success, we verify empirically that, conditional on the expected composition of the committee, its actual composition is uncorrelated with any observable predetermined factor. We estimate equation (3) using predetermined characteristics included in  $\mathbf{X}_i$  as outcome variables instead of controls. As expected, the evidence is consistent with the assignment being indeed random. Table 2 shows estimation results for the eleven predetermined variables that are common for Italian and Spanish databases. Out of forty four coefficients, only two are significantly different from zero at 5 percent level. A joint F-test cannot reject that quality of female and male candidates is similar across committees with different gender compositions.

<sup>20</sup>In Italy, a random sequence of numbers was drawn and was then applied to several disciplines. In Spain, the random draw was carried out publicly on the same day for all disciplines and was certified by the notary.

TABLE 2—RANDOMIZATION CHECK

	1	2	3	4	5	6	7	8	9	10	11
Dependent variable:	Publications										
	All	Articles	Books	Chapters	Patents	Total AIS	A-journal articles	Coauthors per article	Proportion first-author	Proportion last-author	Age
	Italy										
Share of women in committee	0.014 (0.034)	0.005 (0.031)	-0.023 (0.031)	-0.020 (0.028)	0.019 (0.021)	-0.005 (0.030)	0.038 (0.027)	0.017 (0.031)	-0.040 (0.023)	-0.011 (0.041)	-0.065 (0.035)
Female candidate*Share of women in committee	-0.027 (0.079)	-0.001 (0.071)	0.059 (0.066)	0.048 (0.063)	-0.040 (0.049)	0.018 (0.064)	-0.087 (0.061)	-0.044 (0.067)	0.093 (0.053)	0.031 (0.088)	0.150 (0.080)
	Spain										
Share of women in committee	-0.019 (0.030)	-0.029 (0.031)	0.004 (0.022)	0.002 (0.022)	0.024 (0.015)	-0.068 (0.027)	-0.023 (0.022)	-0.040 (0.030)	-0.020 (0.031)	-0.023 (0.032)	0.034 (0.034)
Female candidate*Share of women in committee	0.015 (0.077)	0.038 (0.078)	-0.017 (0.054)	-0.010 (0.055)	-0.062 (0.038)	0.152 (0.068)	0.043 (0.057)	0.103 (0.076)	0.045 (0.078)	0.042 (0.080)	-0.093 (0.086)

*Note:* OLS estimates. All regressions include also the variables *Female candidate*, *Expected share of women in committee*, and the interaction between the two. Standard errors are clustered by committee.

We examine the causal impact of committees' gender composition in column 3 of Table 1. In Italy, the proportion of women in committees has no significant impact on the success rate of male candidates and it has a significant negative impact on the relative chances of success of female candidates (upper panel). An additional female evaluator decreases the relative chances of success of female candidates by approximately 1.8 p.p. ( $\beta_2 = -0.092$ ,  $\Delta Female_e=1/5$ ). In Spain, the share of female evaluators has a positive effect on the success rate of male candidates and a negative effect on the success rate of female candidates, though these effects are not significantly different from zero (lower panel).

To make estimates from Spain and Italy more comparable, it is useful to consider explicitly the upper and the lower bounds of a 95 percent confidence interval. In Italy, an additional woman in the committee decreases the success rate of female candidates relative to men by somewhere between 0.4 and 3.3 p.p. In Spain, an extra woman on the committee can lower it by at maximum 1.0 p.p. but she can also increase it by up to 0.5 p.p. In sum, the impact that women in committees have upon the relative success rate of female candidates is negative and statistically significant only in the Italian case, but we cannot reject that the effect is statistically similar in the two countries.

#### INSTRUMENTAL VARIABLES ESTIMATES

To account for the resignation of some evaluators before the actual evaluation took place, we instrument the final gender composition of the committee using as an instrument the initial composition determined by the random draw. Specifically, we estimate the following equation using the instrumental variables (IV) method:

$$(4) \quad Y_{ie} = \beta_0 + \beta_1 Female_i + \beta_2 Female_e^{final} + \beta_3 Female_i * Female_e^{final} + \beta_4 Female_e^{expected} + \beta_5 Female_i * Female_e^{expected} + \mathbf{X}_i \beta_6 + \epsilon_{ie}$$

where  $Female_e^{final}$  represents the share of female evaluators in the committee that evaluated candidates, and  $Female_e^{final}$  and  $Female_i * Female_e^{final}$  are instrumented using  $Female_e^{initial}$  and  $Female_i * Female_e^{initial}$ .

The first stage results of the IV estimation show that there is a strong relationship between the initial and the final gender composition of committees (see Table D1). The IV estimates are slightly larger but very similar to the intention-to-treat estimates (column 4 of Table 1). To further increase the precision of these estimates, we also reestimate equation (4) including evaluation fixed effects. The estimates are slightly more accurate but they are (statistically) unchanged (column 5 of Table 1).

Female and male evaluators differ in a number of dimensions. As shown in Table C1, male evaluators tend to be relatively older, have longer tenure, and a longer publication record. They are also more likely to be based in the south of Italy and Spain. In order to check whether our results can be explained by these

differences, we estimate equation (4) including the interaction between evaluators' characteristics and candidates' gender. The inclusion of these controls does not affect our previous estimates (Table 1, column 6).

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

*C. Does the presence of women in the committee affect candidates' decision to withdraw?*

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

Therefore, the above estimates may in principle capture the effect that the gender composition of a committee has upon candidates' decision to self-select into the process. To examine this issue, we use data from Italy and estimate equation (4) using as the dependent variable the indicator for those candidates who did not withdraw their application. While relatively fewer women decided to go ahead with the application (-2.6 p.p.), these differences are not related to the share of female evaluators (Table 1, column 7). The evidence thus suggests that committees' gender composition does not affect application decisions and its impact on the chances of success of candidates can be attributed to evaluations.

*D. Does the presence of women in the committee affect the quality of promoted candidates?*

An additional justification for increasing female representation in committees might be that female researchers help to reduce evaluation biases and select better candidates, even though not necessarily more female candidates. To learn about the quality of the assessments, we compare the observable productivity of candidates who qualified in committees with different gender compositions:

$$(5) \quad q_{ie} = \beta_0 + \beta_1 Female_e^{final} + \beta_2 Female_e^{expected} + \epsilon_{ie}$$

where  $q_{ie}$  is a proxy of candidate  $i$ 's quality, measured at the time of the evaluation or during the following five years. We estimate equation (5) for all qualified candidates, and then separately for females and males. We instrument the final gender composition of the committee ( $Female_e^{final}$ ) using the original one ( $Female_e^{initial}$ ).

We consider several proxies of quality. First, we consider the research output of successful candidates at the time of the evaluation. As shown in Table 3, candidates that were promoted by committees with a different gender composition

are at the time of the evaluation statistically similar in terms of the number of papers that they have published, the quality of the journals, the number of students advised or their participation in theses committees.

TABLE 3—QUALITY OF QUALIFIED CANDIDATES

<i>Dep. var.:</i>	1	2	3	4	5	6	7
	Publications	Citations	Total AIS	A-journal articles	PhD students advised	PhD thesis committees	Success in future evaluations
A. Italy, before the evaluation							
All	0.017 (0.088)	0.130 (0.117)	-0.055 (0.157)	-0.135 (0.255)			
Women	-0.044 (0.112)	0.139 (0.143)	0.154 (0.170)	-0.102 (0.317)			
Men	0.029 (0.101)	0.098 (0.150)	-0.208 (0.211)	-0.213 (0.251)			
B. Spain, before the evaluation							
All	0.022 (0.145)	0.072 (0.223)	-0.088 (0.244)	-0.200 (0.237)	0.125 (0.136)	-0.147 (0.132)	
Women	0.210 (0.206)	0.469 (0.370)	-0.004 (0.399)	-0.142 (0.329)	0.580 (0.229)	0.053 (0.220)	
Men	-0.124 (0.193)	-0.242 (0.291)	-0.215 (0.301)	-0.219 (0.333)	-0.170 (0.176)	-0.303 (0.168)	
C. Spain, after the evaluation							
All	0.016 (0.132)	-0.060 (0.218)	-0.098 (0.227)	-0.173 (0.181)	0.175 (0.135)	-0.086 (0.136)	0.042 (0.052)
Women	0.345 (0.213)	-0.009 (0.356)	-0.102 (0.376)	0.170 (0.288)	0.119 (0.212)	-0.117 (0.231)	0.001 (0.054)
Men	-0.187 (0.182)	-0.140 (0.281)	-0.247 (0.284)	-0.266 (0.252)	0.080 (0.191)	-0.134 (0.186)	0.019 (0.077)

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

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

*E. Individual voting*

We have documented that mixed-gender committees are not more favorable towards female candidates than all-male committees. This finding is consistent with several possibilities. It might be that female evaluators are not more favorable towards female candidates than their male counterparts. Alternatively, maybe female evaluators are more sympathetic towards female candidates (or less unbiased) but their presence in the committee induces male evaluators to become less favorable towards female candidates. To shed light on this issue, we analyze the information provided by individual voting reports, available in Italy.

First, we compare the assessments of male and female evaluators sitting in the same committee. We estimate the following equation:

$$(6) \quad V_{ije} = \beta_0 + \beta_1 Female_j + \beta_2 Female_i * Female_j + \mu_{ie} + \epsilon_{ije},$$

where  $V_{ije}$  takes value one if evaluator  $j$  casted a positive vote for candidate  $i$  in evaluation  $e$ , and  $Female_i$  and  $Female_j$  are indicators that capture the gender of the candidate and the evaluator respectively. A vector of application fixed effects  $\mu_{ie}$  captures any differences in application characteristics that are observable to all evaluators.

The empirical results suggest that, if anything, female evaluators are more favorable towards female candidates than male evaluators. Female candidates are 0.7 p.p. (1.6 percent) more likely to receive a positive vote from a female evaluator than from a male evaluator, although this difference is not statistically different from zero (Table 4, column 1). This estimate is likely to be a lower bound of the overall effect. Committee members share information and discuss their decision before casting their vote. A high fraction of committees reach unanimous decisions, suggesting that there may be less disagreement reflected in these final individual evaluations than there would have been at interim stages.

Another question that we would like to answer is whether the voting behavior of male evaluators changes when there are women on the committee. We estimate the following equation on the sample of assessments granted by male evaluators:

$$(7) \quad V_{ije} = \beta_0 + \beta_1 Female_i + \beta_2 Female_{je}^{final} + \beta_3 Female_i * Female_{je}^{final} + \\ + \beta_4 Female_{je}^{expected} + \beta_5 Female_i * Female_{je}^{expected} + \mathbf{X}_i \beta_4 + \epsilon_{ij},$$

where  $Female_{je}^{final}$  and  $Female_{je}^{expected}$  stand respectively for the actual and the expected share of women in a committee including evaluator  $j$ .<sup>21</sup> Coefficient  $\beta_2$  captures how the probability that a male candidate receives a positive vote from a male evaluator varies depending on the gender composition of the committee. Similarly, coefficient  $\beta_3$  captures how the presence of women in the committee

<sup>21</sup>We compute these expectations separately for each evaluator using the outcomes of 1,000,000 simulated random draws that take into account the rules of the randomization.

TABLE 4—INDIVIDUAL VOTING

	1	2	3	4
	All evaluators	Male evaluators		
Female candidate	-	-0.0004 (0.009)	0.008 (0.006)	-0.003 (0.006)
Female evaluator	-0.001 (0.007)			
Female candidate * Female evaluator	0.007 (0.005)			
Share of women in committee		0.017 (0.079)	-	-
Female candidate*Share of women in committee		-0.042 (0.043)	-0.061 (0.030)	-0.078 (0.030)
<i>Controls:</i>				
Application FE	Yes			
Expected share of women		Yes	Yes	Yes
Female candidate*Expected share of women		Yes	Yes	Yes
Candidate characteristics		Yes	Yes	Yes
Exam FE			Yes	Yes
Number of observations	294,656	240988	240988	281289

*Note:* OLS estimates. The dependent variable is an indicator that takes value one if the evaluator casted a positive vote for a given candidate. Column 1 includes information from all individual evaluations, columns 2-4 include information only on evaluations by male evaluators. In column 4 we also include applications that were withdrawn after committee composition was announced, imputing a negative assessment to these applications. Candidate characteristics include all predetermined characteristics listed in Table C2. Standard errors are clustered by committee.

affects the probability that a female candidate receives a positive vote from a male evaluator, relative to a male candidate.

There are three possible threats to the consistency of our estimates. First, similarly to the analysis conducted in previous sections, the initial assignment of evaluators to committees should be random. As shown above, this assumption is satisfied. Second, we do not observe the assessments that would have been casted by evaluators who resigned (8 percent of initial evaluators). This might introduce a selection bias if resignations are related to the gender composition of the committee or to evaluators' gender biases. We examine this possibility in Appendix F. We do not find evidence suggesting that resignations are related to gender issues. Third, given that we only observe the evaluations received by candidates who did not withdraw their application (86 percent of applicants), a bias might arise if candidates' withdrawal decision somehow depends on committees' gender composition or gender biases. Our previous analysis shows that the gender composition of committees does not affect application decisions (see section 4.3). As a robustness check, we also consider an additional specification where we impute a negative assessment to every withdrawn application.

According to our estimates, each additional female evaluator in the committee

increases the probability that a male candidate receives a positive vote from a male evaluator by 0.3 p.p. ( $\beta_2 = 0.017$ ,  $\Delta Female_e^{final} = 1/5$ ) and it decreases the probability that a female candidate receives a positive vote, relative to a male candidate, by 0.8 p.p. ( $\beta_3 = -0.042$ ,  $\Delta Female_e^{final} = 1/5$ ), although these estimates are not significantly different from zero (Table 4, column 2). To increase the accuracy of the estimation, we also include evaluation fixed effects. According to this specification, each additional woman in the committee reduces the probability that a female candidate receives a positive vote from a male evaluator by 1.2 p.p. ( $\beta_3 = -0.061$ ,  $\Delta Female_e^{final} = 1/5$ ), relative to the probability that a male candidate receives a positive vote (column 3). This effect is significant at the 5 percent level. The estimate is slightly larger, around 1.6 p.p., if we consider in our analysis also candidates who withdrew their application (column 4).

#### F. Mechanisms

The two large-scale randomized natural experiments provide a clear result: increasing the proportion of women in scientific committees does not increase the success rate of female candidates. The analysis of individual votes within the committee suggests that this is due to two factors. On the one hand, female evaluators are slightly more likely to vote in favor of female candidates than male evaluators, but this effect is not economically or statistically significant. On the other hand, the presence of women in the committee decreases the probability that female candidates receive a positive vote from male evaluators. Next, we analyze these two issues in more detail.

#### WHY ARE WOMEN NOT MORE SUPPORTIVE OF OTHER WOMEN?

The literature has emphasized several theoretical arguments according to which evaluators are expected to favor same-sex candidates. The most prominent ones are the existence of gender segregation across research networks, gender segregation across subfields of research, gender stereotypes and discrimination against women attaining top positions. Next, we provide an in-depth examination of these theories and we try to understand why they do not play a more important role in our data.

#### GENDER SEGREGATION ACROSS RESEARCH NETWORKS

One of the arguments behind gender quotas is the existence of ‘old boy networks’. If professional connections with committee members help to achieve success and, at the same time, these connections are gendered, female candidates might be at a disadvantage when evaluation committees do not include women. The relevance of ‘old boy networks’ depends on three factors: (i) the extent to which networks are gendered, (ii) the likelihood that applicants are evaluated by a member of their network, and (iii) the magnitude of the connection premium.



First, we examine whether research networks in Spain and Italy are gendered. We consider all possible pairs between candidates and potential evaluators within a given field and we analyze whether the probability of being linked varies with their gender:

$$(8) \quad L_{ij} = \beta_0 + \beta_1 Female_i + \beta_2 Female_j + \beta_3 Female_i * Female_j + \mu_e + \epsilon_{ij}$$

where  $L_{ij}$  stands for any of the observable links between candidate  $i$  and eligible evaluator  $j$ .  $Female_i$  and  $Female_j$  are indicators for female candidates and eligible evaluators, and  $\mu_e$  are evaluation fixed effects.

As expected, links tend to be gendered. The  $\beta_3$  estimate is positive and significant in all specifications, indicating that, when a male eligible evaluator is substituted by a woman, female candidates' likelihood of being connected increases relatively more than male candidates' in every dimension (Table 5).

TABLE 5—GENDER SEGREGATION ACROSS RESEARCH NETWORKS AND SUBFIELDS

	Italy		Spain					
	1	2	3	4	5	6	7	8
	Colleague	Coauthor	Same subfield	Colleague	Coauthor	PhD Advisor	PhD committee	Research overlap
Female candidate	0.0026 (0.0004)	0.0007 (0.0003)	0.0209 (0.0060)	-0.0012 (0.0014)	-0.0003 (0.0002)	-0.0001 (0.0001)	-0.0010 (0.0003)	0.0065 (0.0028)
Female evaluator	0.0017 (0.0009)	-0.0015 (0.0004)	-0.0067 (0.0075)	0.0006 (0.0014)	-0.0015 (0.0002)	-0.0013 (0.0002)	-0.0047 (0.0006)	-0.0110 (0.0017)
Female candidate*	0.0029 (0.0007)	0.0022 (0.0005)	0.0133 (0.0045)	0.0043 (0.0016)	0.0010 (0.0002)	0.0005 (0.0002)	0.0013 (0.0005)	0.0042 (0.0022)
Female evaluator	0.0262 (0.0002)	0.0140 (0.0001)	0.5897 (0.0029)	0.0453 (0.0007)	0.0045 (0.0001)	0.0025 (0.0000)	0.0142 (0.0002)	0.1959 (0.0010)
Observations	2,555,839	2,555,839	1,373,825	5,445,067	5,445,067	5,445,067	5,445,067	4,711,621

*Note:* OLS estimates. The number of observations corresponds to the number of possible pairs between candidates and eligible evaluators with nonmissing information in a given exam. In Italy, only evaluators who are based in an Italian university are considered. All regressions include evaluation fixed effects. Standard errors are clustered by field.

In Italy, the likelihood of observing a female professor with the same affiliation as a female candidate is 0.6 p.p. (20 percent) larger than the likelihood of observing a similar link between a female professor and a male candidate.<sup>22</sup> In Spain, female professors are 0.4 p.p. (8 percent) more likely to be in the same institution as a female candidate, relative to the probability of being affiliated to the same institution as a male candidate. Coauthorships are also relatively more likely when individuals share the same sex. In Italy female professors are 0.3 p.p. (23 percent) more likely to coauthor with a female candidate than with a male one; in Spain the premium is equal to 0.1 p.p. (23 percent). Similarly, PhD supervisions and participation in PhD committees are also gendered. Female professors are 0.04 p.p. (33 percent) more likely to have a female advisee and 0.03 p.p. (3 percent) more likely to have participated in the same dissertation committee as a female candidate.

Another relevant factor is whether candidates benefit from the presence of a member of their network in an evaluation committee. Previous work by Zinovyeva and Bagues (2015) and Bagues, Sylos-Labini and Zinovyeva (2015) documents the existence of a substantial connection premium in qualification exams in Spain and Italy. However, while connections in evaluation committees might be useful, they are relatively rare in a context where evaluations are conducted at the national level. For instance, as pointed out in section II.C, the probability that a candidate and an eligible evaluator are colleagues is around 3 percent in Italy and 5 percent in Spain. The probability that they are coauthors is even lower, around 1.4 percent in Italy and 0.4 percent in Spain. In sum, we observe a relative large degree of gender segregation across networks and also a substantial connection premium, but the impact of these two factors is likely to be attenuated by the scarcity of connections in committees.<sup>23</sup>

Next, we study whether taking into account connections between candidates and evaluators affects our estimates of the impact of committees' gender composition. We estimate the following equation:

$$(9) \quad Y_{ie} = \beta_0 + \beta_1 Female_i + \beta_2 Female_i * Female_e^{final} + \mathbf{L}_{ie}^{final} \beta_3 + \beta_4 Female_i * Female_e^{expected} + \mathbf{L}_{ie}^{expected} \beta_5 + \mathbf{X}_i \beta_6 + \mu_e + \epsilon_{ie}$$

where  $\mathbf{L}_{ie}^{final}$  is a vector including the different types of links between committee members and candidates. We also include as controls the expected proportion of links in the committee  $\mathbf{L}_{ie}^{expected}$  and we instrument the final composition of the committee ( $Female_e^{final}$ ,  $\mathbf{L}_{ie}^{final}$ ) using the outcome of the initial lottery

<sup>22</sup>We have calculated this figure using the information reported in Table 5, column 1. The probability that a female professor and a female candidate in the Italian sample are affiliated to the same university is equal to 3.34 percent (0.0026+0.0017+0.0029+0.0262), and the probability that a female professor and a male candidate are colleagues is equal to 2.79 percent (0.0017+0.0262).

<sup>23</sup>There are may be also weaker links between candidates and evaluators, such as the existence of a common coauthor. Zinovyeva and Bagues (2015) show that these indirect links tend also to be gendered but they do not have a significant impact on evaluation outcomes.

draw ( $Female_e^{initial}$ ,  $\mathbf{L}_{ie}^{initial}$ ). The vector of coefficients  $\beta_3$  provides information about the causal impact of connections in the committee.

Table 6 reports the results of this analysis. In line with the findings of Zinovyeva and Bagues (2015) and Bagues, Sylos-Labini and Zinovyeva (2015), we find that connections with evaluators are helpful for promotion. The presence of a colleague in the committee increases the success rate of connected candidates by 3.6 p.p. (10 percent) in Italy and by 4.6 p.p. (41 percent) in Spain.<sup>24</sup> The impact of coauthors is larger: 4.7 p.p. (13 percent) in Italy and 12.8 p.p. (112 percent) in Spain. Candidates with an advisor in the evaluation committee also enjoy a premium of 9.0 p.p. (79 percent) and when an evaluator has interacted previously with the candidate in some thesis committee the premium is around 2.5 p.p. (22 percent). However, the inclusion of connections as controls in the analysis does not affect significantly our estimates of the effect of evaluators' gender on candidates' success rate (columns 1 and 5 vs. columns 2 and 6). As pointed out above, a plausible explanation for why connections, while being gendered, do not affect significantly our estimates may be related to their scarcity. For instance, in Italy the probability that a female candidate and a male evaluator are coauthors is around 1.4 percent. This probability increases by 0.1 p.p. when the evaluator is also female. Taking into account the premium associated to the presence of a coauthor in the committee (4.7 p.p.), replacing a male evaluator by a female one translates into an increase in the average success rate of female candidates by a mere 0.005 p.p. Moreover, as we show in Appendix G, evaluators' support of connected candidates does not depend on their gender.

<sup>24</sup>To calculate these figures we take into account the number of committee members in Italy and Spain (5 and 7 respectively) and the average success rate in each country (37 percent and 11 percent). For instance, in Italy the presence of a colleague in the committee has an impact of 3.6 p.p. ( $\beta_3 = 0.181$ ,  $\Delta Female_e^{final} = 1/5$ ). Relative to an average success rate of 37 percent, this implies a 10 percent premium.

TABLE 6—CONNECTIONS AND RESEARCH SIMILARITY

	1	2	3	4	5	6	7	8
	Italy				Spain			
Female candidate	0.008 (0.008)	0.006 (0.007)	-0.008 (0.009)	-0.010 (0.009)	-0.011 (0.007)	-0.010 (0.007)	-0.011 (0.008)	-0.011 (0.008)
Female candidate * Share of female evaluators	-0.128 (0.035)	-0.124 (0.035)	-0.061 (0.046)	-0.060 (0.046)	-0.016 (0.028)	-0.020 (0.028)	-0.017 (0.035)	-0.021 (0.035)
<i>Connections in committee:</i>								
Colleagues		0.181 (0.036)		0.180 (0.044)		0.319 (0.031)		0.319 (0.031)
Coauthors		0.237 (0.048)		0.201 (0.053)		0.869 (0.140)		0.840 (0.142)
PhD advisors						0.633 (0.107)		0.575 (0.115)
PhD thesis committee						0.174 (0.037)		0.166 (0.038)
<i>Research similarity:</i>								
Same subfield				0.046 (0.032)				
Overlap in research interests								0.124 (0.037)
<i>Controls:</i>								
Expected connections		Yes		Yes		Yes		Yes
Expected same subfield				Yes				Yes
Expected overlap in research interests				Yes				Yes
Number of observations	69020	69020	35832	35832	31243	31243	27998	27998

*Note:* IV estimates. All regressions include exam fixed-effects, an interaction between *Female candidate* and the *Expected share of women in committee*, and controls for all individual predetermined characteristics listed in Table C2. Connection variables are measured in shares. *PhD thesis committee* refers to candidates and evaluators who have been members of the same doctoral thesis committee. *Same subfield* is the share of evaluators who belong to the same subfield (*settore scientifico disciplinario*) as the candidate. *Overlap in research interests* is based on evaluators' and candidates' participation in doctoral thesis committees, which are classified in 2,000 different subfields (see more details in Data section). *Expected connections* is a vector including the expected share in the committee of *colleagues*, *coauthors*, *advisors* and *PhD thesis committee*. Standard errors are clustered by committee.

## GENDER SEGREGATION ACROSS RESEARCH SUBFIELDS

Another argument in favor of increasing the share of women in committees has been the potential existence of gender segregation across subfields. If committee members tend to prefer candidates with similar research interests and, at the same time, men and women are segregated across research subfields, the lack of women in committees might hinder the ability of female candidates to succeed.

The extent of gender segregation across subfields is likely to depend on the level of aggregation at which evaluations are held. Segregation is probably larger when applicants are grouped in a few broadly defined fields. In the nation-wide evaluations that we analyze in this paper, applicants were classified in approximately 200 different fields (e.g. *Applied Economics*). We check whether, at this level of aggregation, candidates are more likely to have the same research interests as eligible evaluators of the same gender. We estimate equation (8) using as the dependent variable the research similarity between candidates and eligible evaluators. We observe gender segregation across research subfields in both countries but its magnitude is relatively small. In Italy, a female eligible evaluator is 1.3 p.p. relatively more likely to be in the same subfield as a female candidate than in the subfield of a male candidate. In Spain, the overlap between a female eligible evaluator and a female candidate is 0.4 p.p. larger (Table 5, columns 3 and 8).

Research similarity with evaluators tends to increase candidates' chances of success, but the effect of female evaluators on female candidates' relative success rate is unchanged when we control in the estimation for research similarity (Table 6, columns 3-4 and 7-8). This is consistent with the relatively small level of gender segregation observed. In sum, gender segregation across research interests is too limited for female candidates to benefit significantly from more female evaluators in the committee.

## STEREOTYPES

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

Stereotyping might be stronger when evaluators are less informed about candidates' quality. Given that it might be particularly difficult to assess the quality of candidates who do research in subfields that lie far away from evaluators' knowledge, we divide evaluations in two groups based on the distance between evaluators' and candidates' research interests. The evidence suggests that information asymmetries matter, but the presence of women in the committee does not contribute to eliminate potential gender biases. When candidates and evaluators work in similar areas, evaluators' gender does not have a significant impact (Table 7, first row). However, when candidates do research in a different subfield,

female candidates tend to perform significantly worse when there are relatively more women in the committee. This pattern is observed in both countries.

TABLE 7—HETEROGENEITY ANALYSIS

	1	2	3	4
	Italy		Spain	
	$\geq$ median	$<$ median	$\geq$ median	$<$ median
Research overlap	0.011 (0.046)	-0.179 (0.066)	0.081 (0.047)	-0.125 (0.044)
Discipline	SSH	STEMM	SSH	STEMM
	-0.119 (0.058)	-0.133 (0.037)	-0.027 (0.039)	0.003 (0.041)
Feminization of field	$\geq$ median	$<$ median	$\geq$ median	$<$ median
	-0.149 (0.042)	-0.072 (0.057)	-0.018 (0.040)	-0.016 (0.037)
Level of promotion	FP	AP	FP	AP
	-0.111 (0.059)	-0.138 (0.038)	0.120 (0.054)	-0.072 (0.032)

*Note:* IV estimates. The dependent variable is a dummy variable that takes value one if the candidate qualified. Each coefficient corresponds to an independent regression for the corresponding sample. *Research overlap* is a proportion of committee members with similar research interest as defined in section II.D. SSH stands for social sciences and humanities, and STEMM for science, technology, engineering, mathematics, medicine, and psychology. *Feminization of the field* is measured by the proportion of women among full professors in the discipline. FP and AP stand, respectively, for full and associate professors. Standard errors are clustered by committee.

It is also sometimes argued that stereotyping against women is stronger in sciences and mathematics-related disciplines (Reuben, Sapienza and Zingales, 2014). We compare the effect of female evaluators in STEMM and SSH disciplines, but we do not observe any significant differences between these two groups neither in Spain nor in Italy (Table 7, second row).

One might also expect prejudices against women to be stronger in disciplines that are less feminized and, therefore, offer fewer chances to interact with female researchers. We examine separately disciplines with a relatively low and a relatively high proportion of women among full professors. We do not find any evidence suggesting that evaluators in these two groups differ in terms of their preference for candidates of the same sex (Table 7, third row).<sup>25</sup>

#### HIGH-LEVEL POSITIONS

The impact of committees' gender composition might also depend on the importance of the position at stake. Some male evaluators might be reluctant to

<sup>25</sup>In Table H1 in Appendix H we report results from an alternative specification of heterogeneity tests. Instead of splitting the sample in two groups based on the overlap of candidates' and evaluators' research interests and on the degree of feminization of the discipline, we estimate a model with triple interactions exploiting the full range of possible values of these variables. Results from these alternative specifications are in line with the findings discussed in this section.

see a female colleague at the top of the academic career ladder. They might hold negative stereotypes of women, for instance, regarding their leadership or other abilities specific to full professor positions. There might also be a problem of taste-discrimination.

We examine separately the effect of female presence upon the evaluation committee for candidates to full and associate professor positions (Table 7, fourth row). We do not observe any significant differences between these groups of evaluations in Italy, but we do observe a significant difference between exams for full and associate professorships in Spain. Specifically, it appears that in Spain, in committees assessing candidates to full professor positions, a higher female presence has a positive impact on female candidates' relative chances of success. However, the opposite is true in evaluations for promotion to more junior positions.

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

#### ANALYSIS BY DISCIPLINARY GROUPS

Beyond these theories, it might be that the gender composition of committees matters in some specific fields. The previous empirical literature of evaluators' gender does not provide a clear pattern. Two articles that study the role evaluators' gender in Science and Economics find that evaluators tend to prefer candidates of the same sex (Casadevall and Handelsman, 2014, De Paola and Scoppa, 2015), but in two other studies conducted in the same disciplines evaluators exhibit a preference for candidates of the other sex (Broder, 1993, Ellemers et al., 2004). Six other articles in different fields do not find any significant relationship.<sup>26</sup>

Following the official classification of disciplines adopted by the Italian Ministry, we consider 16 different groups of disciplines: Industrial Engineering, Civil Engineering, Physics, Mathematics, Chemistry, Geology, Biology, Veterinary, Medicine, Psychology, Architecture, Economics and Business, Social Sciences, History, Languages and Law. We estimate equation (4) separately for each group and each country, including evaluation fixed effects and instrumenting the final composition of the committee with the initial one. We report these estimates in Figure 1. Out of 32 coefficients, 28 are not significant, one is significantly positive and three are significantly negative. When we take into account in the calculation of standard errors that we are running multiple regressions using a Bonferroni correction none of the coefficients remains significant. Altogether, it is not possible to reject that the impact is similar to zero in any of the different samples. Similarly, we cannot reject the hypothesis that the effect is similar across different fields.

<sup>26</sup>See more details in Table A1.



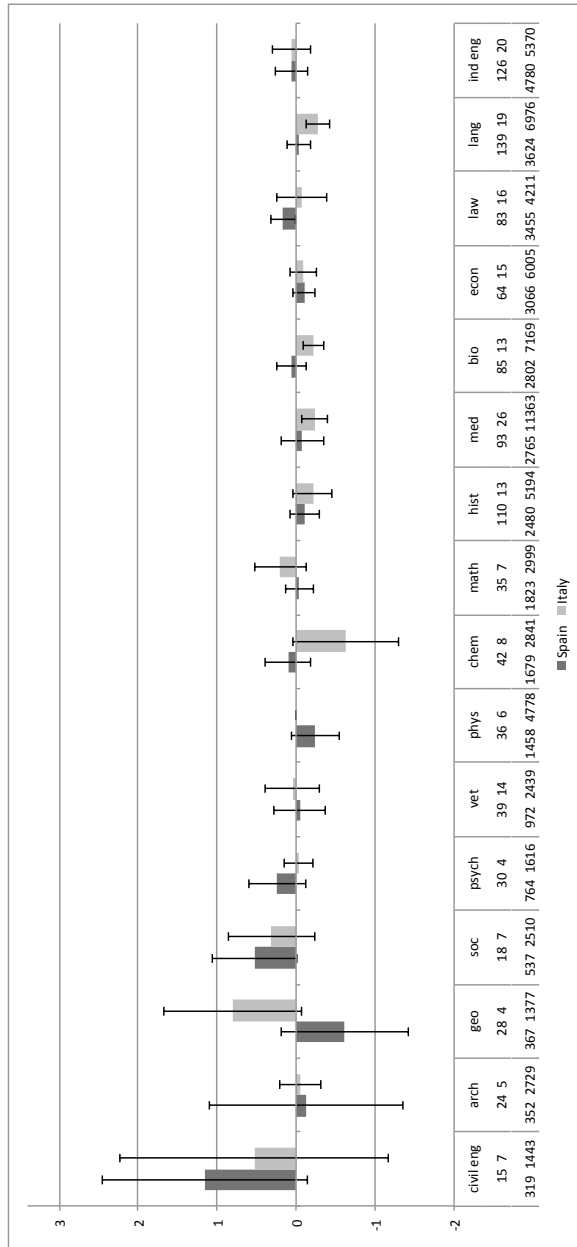


FIGURE 1. THE CAUSAL IMPACT OF COMMITTEES' GENDER COMPOSITION, BY DISCIPLINARY GROUP.

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

WHY DOES THE PRESENCE OF WOMEN IN THE COMMITTEE AFFECT THE VOTING  
BEHAVIOR OF MALE EVALUATORS?

There are at least three potential explanations. The presence of women in the committee might unleash a backlash against female candidates, particularly in fields that have been historically dominated by men (Crocker and McGraw, 1984). While we cannot directly test this hypothesis, we do not observe any significant difference in the impact of committees' gender composition depending on the degree of feminization of the field (see Table 7, third row).

The presence of female evaluators might also induce a licensing effect (Monin and Miller, 2001). In all-male committees, evaluators may feel that they have a moral obligation to worry about sexism and seek to overcome it by expressing more positive (and perhaps less discriminatory) views about female candidates. When there are women on a committee, men may feel licensed to express more honest opinions about female candidates. Furthermore, female evaluators might strengthen male identities within committees and hence weaken their support for female candidates (Akerlof and Kranton, 2000). Unfortunately, we cannot disentangle empirically these two competing hypotheses, licensing effect and male identity priming.

#### IV. Conclusions

A larger presence of women in scientific committees is frequently defended in policy discussions. This paper contributes to this debate by providing a comprehensive and systematic analysis of the impact of scientific committees' gender composition. We exploit the exceptional evidence provided by qualification evaluations for full and associate professorships in every discipline in two different countries, Italy and Spain. These evaluations involved around 100,000 applications and 8,000 evaluators in all academic fields. The random assignment of evaluators to committees creates a setting of large-scale natural randomized experiments. We also take advantage of the availability of very detailed information about candidates, evaluators and the content of evaluations, in order to analyze explicitly the theoretical arguments that are usually employed in support of a higher representation of women in scientific committees.

In general, the presence of female evaluators in the committee neither increases the success rate of female candidates, nor does it alter the quality of selected candidates. Strikingly, in all but one subsamples we observe the opposite pattern in success rates: committees with a higher women share tend to be relatively less favorable towards female candidates. The only exception refers to evaluations to full professorships in Spain, where female candidates have better chances of success when evaluated by a committee with more women.

Information from individual votes within committees suggests that there are two factors that explain why a larger presence of women does not increase the success rate of female candidates. First, while female committee members are

slightly more favorable towards female candidates than their male colleagues, this effect is not economically or statistically significant. Second, male evaluators become less favorable towards female candidates when women are present in the committee, perhaps due to a licensing effect or to male identity priming.

Two common arguments that are usually employed in support of a higher representation of women in scientific committees - gendered networks and segregation across subfields - do not play an important role in our data. We document the existence of gender segregation across research networks in both countries. A female candidate is significantly more likely to be connected to a female evaluator, as measured by coauthorships, affiliation, doctoral thesis supervision and participation in theses committees. We also observe that committees tend to favor connected candidates. However, in the nation-wide evaluations that we consider in this paper the likelihood of connections between candidates and evaluators is small and, therefore, the impact of gendered networks on evaluations is very modest. We also find that evaluators have a preference for candidates with similar research interests but the extent of gender segregation within each field is relatively small. As a result, the impact of gender segregation on evaluation outcomes is very limited. Another justification for increasing the presence of women in committees is that male evaluators may hold stereotypes that have a negative effect upon female candidates. In order to explore the potential impact of gender stereotypes, we focus on cases where information asymmetries are expected to be important. Our results indicate that the gender of evaluators only matters when evaluators are not familiar with candidates' research. However, in this case gender-mixed committees are less favorable towards women than all-male committees.

It remains an open question how the specific institutional characteristics of the Italian and the Spanish promotion systems affect the role of committees' gender composition. Overall, we cannot reject that the estimates for both countries are statistically similar, but we observe a significant difference in the behavior of committees evaluating applications to full professor positions. In Italy, a larger presence of men in the committee increases the chances of success of female applicants. On the contrary, in Spain female applicants to full professorships tend to be relatively less successful when evaluated by an all-male committee. It is unclear whether this difference reflects random sampling or whether it captures some institutional or country-specific characteristic.<sup>27</sup>

Our analysis may be relevant for the design of policies aimed at increasing the representation of women in the academic career. Several countries, including Spain, have introduced quotas in scientific committees requiring the presence of a minimum share of male and female evaluators. According to our results, in general, a higher representation of women in scientific committees *per se* does

<sup>27</sup>Some authors have argued that the degree of transparency in an evaluation procedure can affect gender biases (van den Brink, Benschop and Jansen, 2010). Hence, one possible explanation is that the higher level of transparency and public scrutiny of the Italian system deterred male evaluators from discriminating against female applicants to full professor positions.

not increase the number of promoted female candidates, nor does it help candidates who prove to be more productive in the future. Introducing gender quotas indiscriminately might also have unintended consequences. Quotas may be detrimental for senior female researchers, who would have to spend a disproportionate amount of time sitting on committees and, in some cases, for junior ones, whose chances of success may be hindered.

To be sure, gender quotas could be desirable in certain cases. The analysis suggests that the prevalence of gender segregation across subfields might be an important determinant of whether female committee representation is likely to help female candidates. We expect gender segregation to play a more important role when evaluations are held at a more aggregate level than the one considered here.<sup>28</sup> Another important factor is the potential existence of connections between evaluators and candidates. These connections, which tend to be gendered, are likely to be more relevant in committees at the university- or department-level. More empirical work is needed to understand the impact of gender quotas in those contexts. Moreover, there are certain features of gender quotas that are not captured by our analysis. Evaluators who are explicitly chosen to represent a minority might behave differently, perhaps being more inclined to take a positive view of candidates belonging to their own group. The introduction of quotas may also affect the strategic incentives of evaluators. Nonetheless, keeping in mind these limitations, our results cast doubts on a generalized implementation of gender quotas in scientific committees.

## REFERENCES

- Abe, Yukiko.** 2012. “The Academic Labor Market in Japan and the Presence of Women.” *CSWEP Newsletter*, Fall: 9–10.
- Abrevaya, Jaso, and Daniel S. Hamermesh.** 2012. “Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?” *Review of Economic and Statistics*, 94(1): 202–207.
- Akerlof, George A., and Rachel E. Kranton.** 2000. “Economics And Identity.” *The Quarterly Journal of Economics*, 115(3): 715–753.
- Babcock, Linda, Michele J. Gelfand, Deborah Small, and Heidi Stayn.** 2006. “Propensity To Initiate Negotiations: A New Look At Gender Variation In Negotiation Behavior.” In *Social Psychology and Economics*, ed. David De Cremer, Marcel Zeelenberg and J. Keith Murnighan, 239–259. Mahwah, NY: Lawrence Erlbaum Associates.

<sup>28</sup>The level of disaggregation at which scientific evaluations are held varies largely across countries and institutions. For instance, the European Research Council groups applications in 25 broadly defined areas (<http://erc.europa.eu/evaluation-panels>, accessed on September 1 2015), while in the National Institutes of Health (NIH), which considers only life sciences, grant applications are evaluated by 174 different “study sections” (<http://public.csr.nih.gov/StudySections/Standing/Pages/default.aspx>, accessed on September 1 2015).

- Bagilhole, Barbara.** 2005. "Address at the NORFACE gender equality workshop." *Reykjavik*.
- Bagues, Manuel, and Berta Esteve-Volart.** 2010. "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment." *Review of Economic Studies*, 77(4): 1301–1328.
- Bagues, Manuel, and Maria Jose Perez-Villadoniga.** 2012. "Do Recruiters Prefer Applicants With Similar Skills? Evidence From a Randomized Natural Experiment." *Journal of Economic Behavior and Organization*, 82(1): 12–20.
- Bagues, Manuel, and Maria Jose Perez-Villadoniga.** 2013. "Why Do I Like People Like Me?" *Journal of Economic Theory*, 148(3): 1292–1299.
- Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva.** 2014. "Do Gender Quotas Pass the Test? Evidence from Academic Evaluations in Italy." LEM Working paper 2014/14.
- Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva.** 2015. "Connections in Scientific Committees and Applicants' Self-Selection: Evidence from a Natural Randomized Experiment." IZA Discussion Paper 9594.
- Barres, Ben.** 2006. "Does gender matter?" *Nature*, 442: 133–136.
- Bertrand, Marianne, Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney.** 2014. "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labor Market Outcomes in Norway." NBER Working Paper 20256.
- 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*, 115(501): F81–F107.
- Blau, Francine D., Janet M. Currie, Rachel T. A. Croson, and Donna K. Ginther.** 2010. "Can Mentoring Help Female Assistant Professors? Interim Results from a Randomized Trial." *American Economic Review*, 100(2): 348–352.
- Bloom, Nicholas, Mark Schankerman, and John Van Reenen.** 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica*, 81(4): 1347–1393.
- Bohnet, Iris, Alexandra van Geen, and Max Bazerman.** 2015. "When Performance Trumps Gender Bias: Joint Versus Separate Evaluation." *Management Science*, 62(5): 1225–1234.
- Booth, Alison L., and Andrew Leigh.** 2010. "Do Employers Discriminate by Gender? A Field Experiment in Female-Dominated Occupations." *Economic Letters*, 107(2): 236–238.

- Boschini, Anne, and Anna Sjögren.** 2007. "Is Team Formation Gender Neutral? Evidence from Coauthorship Patterns." *Journal of Labor Economics*, 25(2): 325–365.
- Bosquet, Clément, Pierre-Philippe Combes, and Cecilia Garcia-Peñalosa.** 2013. "Gender and Competition: Evidence from Academic Promotions in France." CESifo Working Paper 4507.
- Brescoll, Victoria L.** 2011. "Who Takes the Floor and Why: Gender, Power, and Volubility in Organizations." *Administrative Science Quarterly*, 56(4): 622–641.
- Broder, Ivy E.** 1993. "Review of NSF Economics Proposals: Gender and Institutional Patterns." *American Economic Review*, 83(4): 964–970.
- Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek.** 2014. "Do Women Shy away from Competition? Do Men Compete too Much?" *Quarterly Journal of Economics*, 129(3): 1409–1447.
- Casadevall, Arturo, and Jo Handelsman.** 2014. "The Presence of Female Conveners Correlates with a Higher Proportion of Female Speakers at Scientific Symposia." *mBio*, 5(1): e00846–13.
- Ceci, Stephen J., and Wendy M. Williams.** 2011. "Understanding Current Causes of Women's Underrepresentation in Science." *Proceedings of the National Academy of Sciences*, 108(8): 3157–3162.
- Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams.** 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science in the Public Interest*, 15(3): 75–141.
- Crocker, Jennifer, and Kathleen M. McGraw.** 1984. "What's Good for the Goose is not Good for the Gander." *American Behavioral Scientist*, 27(3): 357–369.
- De Paola, Maria, and Vincenzo Scoppa.** 2015. "Gender Discrimination and Evaluators' Gender: Evidence from the Italian Academy." *Economica*, 82(325): 162–188.
- De Paola, Maria, Michela Ponzo, and Vincenzo Scoppa.** 2015. "Gender Differences in Attitudes Towards Competition: Evidence from the Italian Scientific Qualification." IZA Discussion Paper 8859.
- Dolado, Juan Jose, Florentino Felgueroso, and Miguel Almunia.** 2012. "Are Men and Women-Economists Evenly Distributed across Research Fields? Some New Empirical Evidence." *SERIEs: Journal of the Spanish Economic Association*, 3(3): 367–393.

- Ellemers, Naomi, Henriette Van den Heuvel, Dick de Gilder, Anne Maass, and Alessandra Bonvini.** 2004. "The Underrepresentation of Women in Science: Differential Commitment or the Queen Bee Syndrome?" *British Journal of Social Psychology*, 43(3): 315–338.
- European Commission.** 2008. "Mapping the Maze: Getting More Women to the Top in Research." Luxembourg: Publications Office of the European Union.
- European Commission.** 2016. "She Figures 2015: Gender in Research and Innovation." Luxembourg: Publications Office of the European Union.
- Fundación Española para la Ciencia y la Tecnología.** 2005. "Mujer y Ciencia: La situación de las Mujeres Investigadoras en el Sistema Español de Ciencia y Tecnología." Madrid: FECYT.
- Ginther, Donna K., and Shulamit Kahn.** 2004. "Women in Economics: Moving Up or Falling Off the Academic Ladder." *Journal of Economic Perspectives*, 18(3): 193–214.
- Ginther, Donna K., and Shulamit Kahn.** 2009. "Does Science Promote Women? Evidence from Academia 1973-2001." In *Science and Engineering Careers in the United States: An Analysis of Markets and Employment.*, ed. Richard B. Freeman and Daniel Goroff, 163–194. University of Chicago Press.
- Hale, Galina, and Tali Regev.** 2014. "Gender Ratios at Top PhD Programs in Economics." *Economics of Education Review*, 41: 55–70.
- Hilmer, Christiana, and Michael Hilmer.** 2007. "Women Helping Women, Men Helping Women? Same-Gender Mentoring, Initial Job Placements, and Early Career Publishing Success for Economics PhDs." *American Economic Review*, 97(2): 422–426.
- Jaffe, Adam B.** 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review*, 76(5): 984–1001.
- Jayasinghe, Upali W., Herbert W. Marsh, and Nigel Bond.** 2003. "A Multilevel Cross-Classified Modelling Approach To Peer Review Of Grant Proposals: The Effects Of Assessor And Researcher Attributes On Assessor Ratings." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 166(3): 279–300.
- Karpowitz, Christopher F., Tali Mendelberg, and Lee Shaker.** 2012. "Gender Inequality in Deliberative Participation." *American Political Science Review*, 106(3): 533–547.
- Khan, Uzma, and Ravi Dhar.** 2006. "Licensing Effect in Consumer Choice." *Journal of Marketing Research*, 43(2): 259–266.

- Kunze, Astrid, and Amalia R. Miller.** 2014. "Women Helping Women? Evidence from Private Sector Data on Workplace Hierarchies." IZA Discussion Paper 8725.
- MECD - Spanish Ministry of Education, Culture and Sports.** 2016. "Estadística de personal de las universidades, Curso 2014-2015." Madrid: MECD.
- Méndez, María J., and John R. Busenbark.** 2015. "Shared Leadership And Gender: All Members Are Equal... But Some More Than Others." *Leadership and Organization Development Journal*, 36(1): 17–34.
- Milkman, Katherine L., Modupe Akinola, and Dolly Chugh.** 2015. "What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway Into Organizations." *Journal of Applied Psychology*, 100(6): 1678–1712.
- MIUR - Italian Ministry of Education, University and Research.** 2016. "Banca dati dei docenti di ruolo. Anno 2014." MIUR: Ufficio di Statistica.
- Monin, Benoît, and Dale T. Miller.** 2001. "Moral Credentials and the Expression of Prejudice." *Journal of Personality and Social Psychology*, 81(1): 33–43.
- Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Graham Mark J., and Jo Handelsman.** 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." *Proceedings of the National Academy of Sciences*, 109(41): 6474–16479.
- 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.
- Niederle, Muriel, and Lise Vesterlund.** 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122(3): 1067–1101.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales.** 2014. "How Stereotypes Impair Women's Careers in Science." *Proceedings of the National Academy of Sciences*, 111(12): 4403–4408.
- Sánchez de Madariaga, Inés, Sara de la Rica, and Juan José Dolado.** 2011. *White Paper on the Position of Women in Science in Spain*. Ministry of Science and Innovation.
- Sandberg, Anna.** 2014. "Competing Biases: Effects of Gender and Nationality in Sports Judging." Stockholm School of Economics. Unpublished.
- Smith, Kristin A., Paola Arlotta, Fiona M. Watt, The Initiative on Women in Science and Engineering Working Group, and Susan L.**



- Solomon.** 2015. "Seven Actionable Strategies for Advancing Women in Science, Engineering, and Medicine." *Cell Stem Cell*, 16(3): 221–224.
- Steinpreis, Rhea E., Katie A. Anders, and Dawn Ritzke.** 1999. "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study." *Sex Roles*, 41(7/8): 509–528.
- van den Brink, Marieke, Yvonne Benschop, and Willy Jansen.** 2010. "Transparency in Academic Recruitment: A Problematic Tool for Gender Equality?" *Organization Studies*, 31(11): 1459–1483.
- Vernos, Isabelle.** 2013. "Research Management: Quotas Are Questionable." *Nature*, 495(7439): 39.
- Vesterlund, Lise, Linda Babcock, and Laurie Weingart.** 2014. "Breaking the Glass Ceiling with "No": Gender Differences in Declining Requests for Non-Promotable Tasks." Carnegie Mellon University. Unpublished.
- Williams, Wendy M., and Stephen J. Ceci.** 2015. "National Hiring Experiments Reveal 2:1 Faculty Preference for Women on Stem Tenure Track." *Proceedings of the National Academy of Sciences*, 112(17): 5360–5365.
- Woolley, Anita Williams, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone.** 2010. "Evidence for a Collective Intelligence Factor in the Performance of Human Groups." *Science*, 330: 686–688.
- Zinovyeva, Natalia, and Manuel Bagues.** 2011. "Does Gender Matter for Academic Promotion? Evidence from a Randomized Natural Experiment." IZA Discussion Paper 5537.
- Zinovyeva, Natalia, and Manuel Bagues.** 2015. "The Role of Connections in Academic Promotions." *American Economic Journal: Applied Economics*, 7(2): 264–292.