Appendix

Data

The data collected include information on 664 candidates for 111 calls posted between 2009 and 2011. The number of calls follows an irregular time profile over the years: 36 calls were posted in 2009, 47 in 2010 and 28 in 2011. The number of applications follows a similar pattern: 183 candidates were examined in 2009, 290 candidates in 2010 and 191 candidates in 2011. Some of these competitions failed to fill the position, while a few resulted in the recruitment of more than one candidate. For 27 competitions, only one application was received and these were excluded from the final sample. Candidates applied online or via email, and their applications are kept by the institution for up to five years. The administrative archives hold the official final reports with the results of each opening as well as the job advertisement contained in the public call.

The final data set used in the analyses comprises 78 calls: 171 candidates for 25 calls posted in 2009, 266 candidates for 31 calls in 2010, and 171 candidates for 22 calls referred to 2011. Out of the 78 calls, 72 resulted in the recruitment of one candidate, and these account for 502 candidates, while five calls led to two recruitments, and one led to three, with 73 and 33 applicants respectively.

Control variables

Age is a continuous variable. The candidate's age is missing in 47% of the CVs. When no age is reported, it is estimated from the year of graduation. Origin was coded distinguishing between region where the institute is placed (local), the rest of Italy, the European Union (including Switzerland) and the rest of the world.

Several variables show the research potential of the candidates. Educational attainment consists of four categories: no degree, bachelor's degree, master's and Ph.D. Since the positions advertised at the institute are research positions that usually require a doctoral degree, a dummy variable Ph.D. (1 if the candidate holds a Ph.D. and 0 otherwise) is used to represent educational attainment in the models.

Years of work experience (research and non-research related activities) are included in the analysis; Ph.D. activity alone is not considered as work experience, whereas lecturing and/or other types of employment, even if undertaken during the Ph.D., are taken into account.

The data set also includes information on the candidates' scientific output. Candidates self-reported their publications in their CVs. Nevertheless, in order to standardise the comparison within and between competitions, the data regarding the number of publications and citations were directly retrieved from Scopus, Elsevier's bibliographic database created in 2004. In order to obtain the applicants' publication records, the Scopus Author search page was queried with the researchers' first and last names. If the author's name was not unique, the results were cross-checked with data appearing on the candidate's CV, such as age, origin and field of study, in order to refine the results and ensure a correct attribution of publications to candidates. The Scopus database was accessed in August 2014. Although other bibliographic sources such as Google Scholar and Web of Science are available, many studies suggest that Scopus is superior in terms of both coverage and accuracy. According to Falagas et al. (2008, p. 338), "Scopus offers about 20% more coverage than Web of Science, whereas Google Scholar offers results of inconsistent accuracy." Moreover, "Scopus helps distinguish between the researchers in a more nuanced fashion than Web of Science." (Meho and Rogers, 2008, p. 1711). Both publications and H-index data were retrieved for the specific year of the call for which the candidate applied. The candidates' H-index was collected in addition to the number of publications, since it combines measures of quantity and visibility of scientific production, and therefore is a good proxy for the importance and significance of candidates' contributions (Hirsch, 2005). However, in order to take into account the different publication propensities across disciplines, we implemented a standardization of H-index within: i) social sciences correlated to engineering research areas (e.g. management, human resources); ii) hard sciences excluding mathematics, iii) engineering, computer sciences, architecture, environmental sciences and mathematics.

The H-index is then adjusted for maternity risk. We obtained data regarding fertility of female researchers from the HR department of the institute. The maternity rate in the institute is very low (only one woman out of four has ever had children), and the average age of mother at the birth of her first child is 35 years old. Adjusted standardized H-index has been computed as follows:

$$H\ index_{std.adj.} = \frac{H\ index_{std.}}{(1 - maternity\ risk)}\ .$$

Table A1 – Variable definitions.

Variable	*Dummy*	*Definition*
Female	D	The candidate is female
Success	D	The candidate is selected for the call
Female (no HR) on committee	D	The candidate is evaluated by a committee with at least one female researcher (no HR)
H-index standardized adjusted		Candidate's H-index at the year of the call, standardized by applicants' field of research and adjusted for maternity risk
Age		Candidate's age in years
Italian local origin	D	The candidate has Italian nationality and local origin (from the region where the institute is located)
Italian non-local origin	D	The candidate has Italian nationality and non-local origin (outside the region where the institute is located)
EU origin	D	The candidate has an EU nationality (including Switzerland)
Non-EU origin	D	The candidate has a non-EU nationality
PhD	D	The candidate holds a PhD
Work experience		Candidate's work experience in years, excluding PhD years
Ties with the institute	D	The candidate has at least one among the types of ties indicated at the "*Candidate-level variables*" subsection
Ties with committee	D	The candidate has ties with the institute either of type 1), 2) or 5) – see main text ("*Candidate-level variables*" subsection)
Ties with institution	D	The candidate has ties with the institute either of type 3) or 4) – see main text ("*Candidate-level variables*" subsection)
Intensity of ties		Number of candidate's types of ties with the institute (range 0-5)
Male-Male ties	D	The candidate is a male and has a tie with a male
Female-Male ties	D	The candidate is a female and has a tie with a male
Undefined	D	The gender of the candidates' contact is not observable
Tenure-track positions	D	R3, R4 positions
Non tenure-track positions	D	T4, post-doc, or co.co.pro. positions
Monthly wage		Monthly gross wage of the posted position expressed in Euros
Contract length		Contract length of the posted position expressed in months

Descriptive statistics

Descriptive statistics regarding the monthly wage, the applicants' age and the contract length across the different research positions are reported in Table A2.

Table A2 – Descriptive statistics of research levels.

Variable	*Mean*	*Std. Dev.*	*Min*	*Max*
Monthly gross wage (all)	2687	905.23	750	4388
R3	3113	124.32	2667	3156
R4	2842	124.22	2750	3150
T4	2750	0	2750	2750
Post-doc	3156	0	3156	3156
Co.co.pro.	2394	815.45	750	4388
Applicant's age (all)	30.65	5.77	19	60
R3	32.94	5.43	26	55
R4	30.65	5.30	24	56
T4	26.59	5.38	20	37
Post-doc	32.33	3.24	26	37
Co.co.pro.	30.17	5.54	19	60
Contract length in months (all)	21.46	11.01	3	44
R3	29.63	9.12	12	42
R4	25.58	9.41	8	36
T4	8.36	1.18	8	12
Post-doc	30.40	6.20	24	36
Co.co.pro.	18.83	10.50	3	44

Table A3 depicts the number of candidates and calls across the different positions included in the data set. There are on average eight applications per call (7.27 for tenure-track positions and 8.06 for non-tenure-track positions). The variation in the number of candidates across calls, however, is considerable, ranging from two applicants to more than thirty applicants. The majority of calls are for non-tenure-track positions (69%). The percentage of female candidates does not differ between tenure-track and non-tenure-track positions. However, the success rate of the female candidates is higher in tenure-track applications.

Table A3 – Number of candidates and calls across different positions.

Position	*N. candidates*	*N. calls*	*Average N. of candidates per call*	*% of female candidates*	*Success rate of female candidates*
Tenure-track	189	26	7.27	0.20	0.22
R3	65	13	5.00		
R4	124	13	9.54		
Non-Tenure-track	419	52	8.06	0.20	0.13
T4	22	2	11.00		
Post-doc	15	2	7.50		
Co.co.pro.	382	48	7.96		
Total	608	78	7.80	0.20	0.20

By breaking down the variable representing prior ties with the institute into "institution" and "committee" ties, it is shown that female candidates have disproportionally fewer ties with the committee than their male counterparts. Similarly, intensity of ties is higher for male candidates (Table A4).

Table A4 – Breaking down of ties with the institute.

	Institution	*Committee*	*Intensity of ties (1-5)*
Females	27.08%	15.38%	1.13
Males	72.92%	84.62%	1.42
Total	100%	100%	1.35

Notes: Candidates with at least one tie amount to 65, of which 15 are females and 50 males.

Table A5 shows that out of 65 observations with prior ties (11% of the final sample), 69.2% of these are male-male direct connections, 20% are female-male connections and 10.8% are undefined.

Table A5 – Gendered ties.

	%
Female-Female tie	0
Female-Male tie	20
Male-Female tie	0
Male-Male tie	69.2
Undefined	10.8
Total	100

Empirical analyses

All models were also estimated replacing the standardized H-index of candidates, adjusted for maternity risk, with the standardized and adjusted number of publications at the year of the competition, without differences in terms of sign and statistical significance of the relevant coefficients. Moreover, the results of the final conditional logit models are mostly confirmed in significance and in signs when linear probability model estimation is performed.

Table A6 reports the full results of a linear probability model, with fixed effects at the call level, which represents the main model specification of Table 4 in the main text.

Table A6 – Probability of success – Linear probability model.

Success	1	2	3	4	5
Female	-0.004	-0.054	-0.046	-0.048	-0.038
	(0.043)	(0.048)	(0.048)	(0.049)	(0.048)
Female × Female (no HR) on Comm.		0.205**	0.197**	0.175*	0.138
		(0.090)	(0.089)	(0.089)	(0.092)
H-index std. adj.			0.037*	0.027	0.021
			(0.021)	(0.019)	(0.020)
Age				-0.005	-0.011
				(0.015)	(0.015)
Age squared				-0.000	-0.000
				(0.000)	(0.000)
Italian (non-local) origin				0.002	0.046
				(0.064)	(0.060)
EU origin				-0.092	-0.049
				(0.063)	(0.059)
Non-EU origin				-0.143**	-0.117**
				(0.062)	(0.058)
PhD				0.109**	0.087*
				(0.045)	(0.044)
Work experience				0.011**	0.012**
				(0.005)	(0.005)
Ties with the institute					0.313***
					(0.083)
R^2	0.135	0.143	0.153	0.190	0.249
Observations	608	608	608	608	608

Notes: The table reports linear probability model coefficients, with call fixed effects. The dependent variable is a dummy variable for being selected. Clustered standard errors (at the call level) are reported in parenthesis. Symbols *, ** and *** indicate that the coefficients are statistically significant at the 10%, 5% and 1% levels, respectively. Italian local origin is used as reference category. The model contains 608 observations in contrast to 606 observations in the conditional logit, which is due to difference in estimation procedures.

Table A7 reports the results of a linear probability model, with fixed effects at the call level, for sample splits as in Table 5 in the main article.

Table A7 – Probability of success in different sample splits – Linear probability model.

Success	1a	1b	2a	2b	3a	3b	4a	4b
	H-index commission:				Position:			
	4th quartile		1st-3rd quartiles		Non-tenure-track		Tenure-track	
Female	-0.015	-0.015	-0.048	-0.036	-0.102**	-0.093*	0.083	0.092
	(0.043)	(0.044)	(0.067)	(0.065)	(0.049)	(0.052)	(0.091)	(0.082)
Female× Female (no HR) on Comm.			0.145	0.100	0.229**	0.189*	0.085	0.054
			(0.106)	(0.109)	(0.098)	(0.104)	(0.186)	(0.196)
H-index std. adj.	0.033	0.033	0.020	0.021	0.052	0.053	0.009	-0.004
	(0.025)	(0.025)	(0.033)	(0.031)	(0.036)	(0.035)	(0.010)	(0.009)
Age	0.004	0.004	-0.008	-0.014	-0.011	-0.017	-0.003	-0.006
	(0.020)	(0.020)	(0.019)	(0.018)	(0.018)	(0.018)	(0.032)	(0.033)
Age squared	-0.000	-0.000	-0.000	0.000	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Italian origin	0.103	0.103	0.133**	0.122**	0.100*	0.099*	0.189***	0.186**
	(0.082)	(0.078)	(0.053)	(0.054)	(0.052)	(0.052)	(0.067)	(0.072)
Ph.D.	0.017	0.017	0.148**	0.104	0.088	0.088	0.140*	0.083
	(0.022)	(0.021)	(0.065)	(0.067)	(0.054)	(0.059)	(0.080)	(0.075)
Work experience	0.014*	0.014*	0.011*	0.012*	0.012*	0.014**	0.009	0.009
	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)
Ties with the institute		-0.001		0.341***		0.263**		0.377***
		(0.235)		(0.093)		(0.112)		(0.130)
R^2	0.160	0.160	0.183	0.257	0.215	0.253	0.170	0.273
Observations	168	168	440	440	419	419	189	189

Notes: The table reports linear probability coefficients, with call fixed effects. The dependent variable is a dummy variable for being selected. Clustered standard errors (at the call level) are reported in parenthesis. In models 1a and 1b the interaction term is not included due to the existence of just one mixed-gender commission belonging to the 4th quartile of commissions in terms of H-index std. adj..Due to small number of cases in sample splits, geographical origin is given as being Italian or not. Symbols *, ** and *** indicate that the coefficients are statistically significant at the 10%, 5% and 1% levels, respectively.

Finally, Table A8 contains estimations from a linear probability model which parallel those of Table 6 in the main text for a conditional logit.

Table A8 – Probability of success – Linear probability model – low vs high wage positions.

Success	Low-wage		High-wage	
Female	-0.101	-0.098	0.004	0.009
	(0.078)	(0.078)	(0.057)	(0.056)
Female × Female (no HR) on Comm.	0.0654	-0.017	0.274**	0.275**
	(0.169)	(0.160)	(0.112)	(0.113)
H-index std.	-0.071	-0.066	0.031	0.028
	(0.064)	(0.048)	(0.020)	(0.020)
Age	-0.023	-0.031	0.017	0.015
	(0.031)	(0.029)	(0.014)	(0.014)
Age squared	-0.000	0.000	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Italian origin	0.164***	0.133**	0.111*	0.120*
	(0.0524)	(0.060)	(0.062)	(0.063)
Ph.D.	0.176*	0.121	0.088*	0.084**
	(0.088)	(0.091)	(0.043)	(0.041)
Work experience	0.019*	0.018*	0.009	0.010*
	(0.011)	(0.010)	(0.005)	(0.005)
Ties with the institute		0.429***		0.128
		(0.119)		(0.107)
R^2		0.331		0.198
Observations		263		345

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Non-Italian origin is used as reference category.

Gender, ties and recruitment

Table A9 reports the percentage of candidates with and without prior ties: overall, 12.40% of the female applicants have previous ties with the institute, and the percentage for males stands at 10.27% (the differences in the proportions are not statistically significant; two-group test of proportions: z = -0. 599, p = 0.549). No statistically significant differences across genders were found when successful and unsuccessful applicants were considered separately.

Table A9 – Gender composition of successful and unsuccessful candidates.

	Females	*%*	*Males*	*%*
Overall				
Ties	15	12.40%	50	10.27%
No ties	106	87.60%	437	89.73%
Successful				
Ties	6	31.58%	22	33.33%
No ties	13	68.42%	44	66.67%
Unsuccessful				
Ties	9	8.82%	28	6.65%
No ties	93	91.18%	393	93.35%

Table A10 reports the predicted probabilities of success for female and male candidates with different profiles, net of other characteristics, computed from estimated coefficients of the main model with two-way and three-way interactions (gender of the candidate, woman on the committee, prior ties). Male candidates have a 40% probability of success without prior ties, and 76% with prior ties. Female candidates with ties and all-male committees have a 55% probability of winning a selection. The probability is very high, at 91%, if female candidates are both connected and there is a female researcher on the selection committee. Moreover, additional analyses show that in 50% of the competitions in which female candidates with a prior connection were successful, there was at least one female researcher on the committee.

Table A10- Conditional logit with 1) interaction between the gender of the candidate and prior ties, and 2) interaction between the gender of the candidate, the presence of women on the committee and prior ties. Predicted probabilities of success.

Candidate gender-prior ties	Average partial effect	Std.error
Male – No ties	0.40	0.90
Male – Ties	0.76	0.70
Female – No ties	0.36	0.86
Female – Ties	0.59	0.91
Candidate gender – Women on committee – Prior ties		
Female – No women on committee – No prior ties	0.45	0.95
Female – No women on committee – With prior ties	0.55	0.95
Female – Women on committee – No prior ties	0.56	0.96
Female – Women on committee – With prior ties	0.91	0.35

Notes: Predicted probabilities are computed from estimated coefficients from Table 4, column 5 and from a model with three-way interactions that is available upon request.

Table A11 reports the results of a linear probability model, with fixed effects at the call level, for the model with types, intensity of ties and gendered ties as in Table 7 in the main text.

Table A11– Probability of success – Linear probability model with types, intensity of ties and gendered ties

Success	1	2	3	4
Female	-0.048	-0.043	-0.026	-0.029
	(0.049)	(0.046)	(0.048)	(0.046)
Female × Female (no HR) on Comm.	0.175*	0.139	0.119	0.141
	(0.089)	(0.087)	(0.089)	(0.087)
Institution ties		0.442***		
		(0.093)		
Committee ties		0.073		
		(0.070)		
Intensity of ties			0.251***	
			(0.045)	
Male-Male ties				0.363***
				(0.103)
Female-Male ties				0.301
				(0.195)
Undefined				0.068
				(0.169)
CONTROLS	YES	YES	YES	YES
R^2	0.190	0.286	0.273	0.255
Observations	608	608	608	608

Notes: The table reports linear probability model coefficients, with call fixed effects. The dependent variable is a dummy variable for being selected. Clustered standard errors (at the call level) are reported in parenthesis. In specification (4) the omitted category for the last three variables is having no ties. The symbols *, ** and *** indicate that the coefficients are statistically significant at the 10%, 5% and 1% levels, respectively. Controls include: H-index std. adj., Age, Age squared, Italian origins, Ph.D., Work experience.

References

Hirsch JE (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16569–16572.

Falagas ME, Pitsouni EI, Malietzis GA and Pappas G (2008). Comparison of PubMed, Scopus, Web of Science, and Google scholar: strengths and weaknesses. *The FASEB Journal*, 22(2):338–342.

Meho LI and Rogers Y (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11):1711–1726.