European Sociological Review, 2020, 1–12
doi: 10.1093/esr/jcaa050
Original Article

Gender Bias in Academic Recruitment? Evidence from a Survey Experiment in the Nordic Region

Magnus Carlsson [1], Henning Finseraas [2], Arnfinn H. Midtbøen [3,*] and Guðbjörg Linda Rafnsdóttir [4]

[1]Department of Economics and Statistics, Linnaeus University, Växjö 35195, Sweden, [2]Department of Sociology and Political Science, Norwegian University of Science and Technology (NTNU), Trondheim 7491, Norway, [3]Institute for Social Research, Oslo 0208, Norway and [4]Faculty of Sociology, Anthropology and Folkloristics, Faculty of Social and Human Sciences, University of Iceland, Reykjavík 102, Iceland

*Corresponding author. Email: ahm@socialresearch.no

Submitted June 2020; revised September 2020; accepted September 2020

Abstract

Gender disparities in top-level academic positions are persistent. However, whether bias in recruitment plays a role in producing these disparities remains unclear. This study examines the role of bias in academic recruitment by conducting a large-scale survey experiment among faculty in Economics, Law, Physics, Political Science, Psychology, and Sociology from universities in Iceland, Norway, and Sweden. The faculty respondents rated CVs of hypothetical candidates—who were randomly assigned either a male or a female name—for a permanent position as an Associate Professor in their discipline. The results show that, despite the underrepresentation of women in all fields, the female candidates were viewed as both more competent and more hireable compared to their male counterparts. Having children or a stronger CV do not change the overall result. Consequently, biased evaluations of equally qualified candidates to Associate Professor positions do not seem to be the key explanation of the persistent gender gap in academia in the Nordic region.

Introduction

Gender disparities in top-level academic positions are slow to change. Despite a remarkable progress in women's educational attainment since the 1960s (van Hek, Kraaykamp and Wolbers, 2016; OECD, 2018), as well as a slow gender convergence in attainment of faculty positions (European Commission, 2019), women still tend to be underrepresented in professor positions (Ceci, 2018; Cech and Blair-Loy, 2019). This holds true even in the Nordic countries, well known for their institutionalized gender equality norms (Teigen and Skjeie, 2017). Although the Nordic countries rank at the top of the Global Gender Gap Index, benchmarking 153 countries on their progress towards gender parity (World Economic Forum, 2020), women are strongly underrepresented in professor positions in the Nordic region (e.g., Bergman and Rustad, 2013; Nielsen, 2017).

To some, the persistent gender gap in top-level academic positions reflects historical differences in educational attainment and academic careers that will naturally evaporate with time, as women eventually will catch up with men (e.g., De Groot, 1997; Allen and

Castleman, 2001). Others point to various sorting mechanisms as the key explanation for persistent gender inequality in academia: Because fewer women embark on an academic career, especially in the fields of science, technology, engineering, and mathematics (STEM), the underrepresentation of female professors in these disciplines might be a matter of choices—constrained or not—made long before they enter the academic system (Ceci and Williams, 2011; Ceci, 2018). The higher representation of women in the social sciences and humanities compared to STEM indeed indicates that sorting is an important mechanism producing gender inequality. However, women tend to be underrepresented at the professor level in social sciences and humanities as well, despite a strong female dominance among students (Santos and Dang Van Phu, 2019), suggesting that the phenomenon of 'leaky pipelines' are present even in these fields.

A range of studies point to various forms of gender bias as a major explanation for the persistent underrepresentation of women at the top of the academic hierarchy. Research has documented that male and female students receive differential access to mentoring (Blau et al., 2010) and entry-level positions as lab assistants (Moss-Racusin et al., 2012). Some studies show that female researchers face bias-driven barriers in peer-review processes (Wennerås and Wold, 1997) and experience more difficulties in achieving promotion and tenure (Barbezat and Hughes, 2005). Research also shows that family responsibilities affect the academic performance of men and women differently (e.g., Lutter and Schröder, 2019), yet whether this is due to bias is a debated question. Anders (2004) suggests that women self-select away from academia in response to barriers related to parenthood, while Xu (2008) claims that women's stronger turnover is due to dissatisfaction with research support, advancement opportunities, and free expression of ideas, rather than gender-based differences in family responsibilities. Supporting the latter view, recent research finds that hiring committees penalize female scholars in relationships because they are considered 'unmovable'—and hence less attractive—compared to single women and men regardless of relationship status (Rivera, 2017).

In terms of gender bias in academic *recruitment*, however, the existing evidence is mixed. To be sure, gender discrimination in academic recruitment has been a topic of study for decades, often suggesting that there is a female advantage in hiring (Baldi, 1995; Merritt and Reskin, 1997; Wolfinger, Mason and Goulden, 2008; National Research Council, 2009; Lutter and Schröder 2016). However, this body of work mainly consists of observational studies. More recently, experimental studies have examined the occurrence of gender bias directly, but, as in the literature on gender discrimination in the labour market in general (Rich, 2014; Neumark, 2018), the findings vary. Experiments involving hiring to non-faculty positions within universities tend to find a bias in favour of male job applicants (Foschi, Lai and Sigerson, 1994; Steinpreis, Anders and Ritzke, 1999; Moss-Racusin *et al.*, 2012), but for higher-level positions in academia, experimental studies of gender bias in evaluations find that female applicants have an advantage over male applicants with similar qualifications (Williams and Ceci, 2015; Ceci, 2018).

This article presents the results of a survey experiment that investigates gender bias in recruitment among faculty ($N = 775$) at universities in Iceland, Norway, and Sweden in six disciplines (Economics, Law, Physics, Political Science, Psychology, and Sociology) where women are underrepresented at the level of Full Professor. In the experiment, the faculty respondents rated CVs of hypothetical candidates—who were randomly assigned either a male or a female name—for a permanent position as an Associate Professor in their discipline. We asked faculty members to rate the candidates according to their competence and hireability. The design allows us to detect the occurrence of possible gender bias in evaluations of male and female candidates to Associate Professor positions. As each participant evaluated two CVs of the same candidate (one of them having a more extensive publication record) and were randomly assigned candidates with two children or no children, we are moreover able to examine the effects of children and publication records on faculty's assessments of competence and hireability, and whether these effects differ for men and women.

Our research design combines the merits of previous studies of gender bias in academic recruitment. We build on Moss-Racusin *et al.*'s (2012) experimental design, asking faculty to rate fictitious applicants' competence and hireability based on actual applicant material (CVs), while similar to Williams and Ceci (2015), we study faculty's evaluations of applicants for permanent associate professor positions rather than early-career or even student assistant positions. We used actual CVs because evaluating a CV is an important part of the hiring process in which stereotypes may play a role, while at the same time being more neutral than a highly positive narrative, which may leave limited room for bias in the evaluation. Also, highly positive narratives, as in Williams and Ceci (2015), may signal very qualified candidates, which could limit the external validity of the results (see Heckman, 1998, for an elaboration of this

point). Thus, we carefully calibrated the CVs against an existing pool of real CVs and pilot tested them to make them representative for the respective discipline. Each participant evaluated only female or male applicants to avoid that they realized that the purpose of the experiment was to study gender disparities (and perhaps provided socially desirable answers, e.g., rated the female applicant higher).

Existing experimental studies on gender bias in academic recruitment are conducted exclusively in the United States and focus only on the STEM fields, which prevents the results from being generalized to the European context and to other fields of study. Hence, the current study is the first of its kind to examine whether bias in recruitment is a likely explanation for the prevailing gender gap in top-level academic positions in Europe.

We continue by presenting the theoretical framework and the preregistered hypotheses of the study. Next, we detail the research design before presenting the main results. In conclusion, we interpret our findings and discuss their implications for the broader understanding of persistent gender inequality in academia.

Theoretical Framework and Hypotheses

Most theories in the social sciences would suggest that there is a bias against women in academic recruitment. The stereotype content model, for example, is a model of cognition that has received growing support for explaining bias between groups of people in society, including in the labour market (Fiske et al., 2002; Abele *et al.*, 2008). In this model, communion and agency are central dimensions of social judgements that are assumed to stem from evolution. Research in social cognition has established that people judge other people on the basis of these dimensions (Judd et al., 2005; Fiske, Cuddy and Glick, 2007). The communion dimension concerns people's intentions and involves judgements of other's intent to either harm or help. The agency dimension involves judgements of other people's capacity to carry out the intent. People high on communion tend to be characterized as trustworthy, empathic, and friendly, while people high on agency tend to be characterized as intelligent, skilled, creative, and efficient.

When people rate individuals on these dimensions, women tend to receive higher scores on communion than men, while men tend to receive higher scores on agency (e.g., Fiske and Dupree, 2014). Apparently, such gender stereotypes also exist for more specific groups of people, such as professors. In line with the stereotype content model, previous research has shown that female professors are perceived as less qualified and have their work less valued compared to males, but are instead perceived to possess feminine traits that include warmth and accessibility (Miller and Chamberlin, 2000; Monroe et al., 2008).

An implication of the stereotype content model is that a gender bias in hiring situations can emerge if there are gender stereotypes about the traits that men and women possess and these traits are incompatible with the traits perceived to be needed for the job (Fiske et al., 1991; Burgess and Borgida, 1999). The question is then what traits are perceived to be needed to work as an Associate Professor? Fiske and Dupree (2014) find that the professor occupation is perceived to require more of typical male traits (the agency dimension) and less of female traits (the communion dimension). To the extent that these findings apply also for Associate Professor positions, they suggest a potential mismatch between the perceived traits that female applicants possess and the perceived traits needed in the occupation. Together with the empirical fact that women are underrepresented in top-level academic positions, this motivated the following main hypothesis in the survey experiment:

H1: Female applicants have lower ratings on competence and hireability than males.

Of course, the main hypothesis could also be motivated from theories of discrimination in other disciplines, e.g., economics and sociology, although the mechanisms might look slightly different. In economics, theories of statistical discrimination suggest that a gender bias in hiring situations could emerge because of incomplete information about the productivity of the applicants (Phelps, 1972; Arrow, 1973). In our case, men could, on average, be perceived as having greater learned abilities and being more productive in the professor occupation, which could be a result of historical gender imbalances and biases. It is also possible that women are expected to be less productive in the professor occupation because of a higher likelihood of having career disruptions due to greater family responsibilities and a perception that long-term parental leave could risk the progress of ongoing research projects. These cases belong to the type of statistical discrimination in which a prospective employer classify applicants on the basis of group belonging and make judgements based on the groups average productivity rather than on the individual's productivity which is partly unobserved.

Taking a sociological perspective, cultural beliefs about the gendered nature of jobs could also lead to bias in evaluations (e.g., Reskin and Roos, 1990; Ridgeway and England, 2007). Indeed, much research has shown that gender segregation in the labour market is not only

a product of gender differences in human capital, but also a consequence of how women and men are viewed as suited for different jobs and work tasks (e.g., Correll, 2004; Correll, Benard and Paik, 2007; Ridgeway, 2011; Orupabo, 2018). In order to understand the sources of such social inequality, expectation states theory (e.g., Correll and Ridgeway, 2003; Ridgeway and Bourg, 2004) suggests that hierarchies of evaluation and influence are based on gendered beliefs about social status, which produce different expectations as to what men and women can accomplish. Following this logic, we would expect that members of academic hiring committees could—consciously or not—translate ideas about the gender of the applicant into discriminatory behaviour against women based on gender categorization and stereotyped ideas of skills and suitability.

We further formulated an inferior hypothesis, which concerns gender differences in the perceptions of applicants with and without children. The idea is that having children could be an important moderator for a potential gender bias in hiring. Cuddy, Fiske and Glick (2004) show that working women who become mothers receive higher ratings on communion, while their ratings on agency decrease. Thus, women seem to trade perceived agency for perceived communion. However, working men who become fathers gain in perceived communion and maintain perceived agency. If this finding applies also for applicants for Associate Professor positions, the mismatch between the traits that women possess and the traits perceived to be needed for the job should increase. Thus, we expect a greater gender bias in favour of men among applicants with children. This motivates the following hypothesis:

H2: Females have a lower return to children than males.

Finally, we formulated an additional inferior hypothesis concerning the effect of having a strong CV. As we explain below, to help conceal that the main intent of the experiment was to evaluate the prevalence of gender bias, and thereby avoid socially desirable answers, we varied the quality of the CVs. Thus, mainly as a byproduct, we were able to investigate the quality of the CV as a possible moderator for the gender bias. A certain type of statistical discrimination models would predict that females have a lower return to a strong CV. In particular, this type of model may apply in environments where female applicants are uncommon, such as in male-dominated disciplines in academia and/or in environments in which there are few female professors. The idea is that in this context a CV could be a noisier productivity signal for female applicants than for male applicants. As a result, there could be a gender bias in favour of men, even if men and women have the same perceived average productivity. We formulated the following hypothesis:

H3: Females have a lower return to a strong CV than males.

Research Design

In the experiment, the faculty respondents rated CVs of hypothetical candidates for a permanent position as an Associate Professor in their discipline. The experimental design closely resembles academic appointments in the Nordic region, where hiring procedures for academic staff in public higher education in these countries are strictly regulated. Vacant posts are normally advertised openly and internationally. Applications to vacant positions typically include an application letter, a CV, and the applicants' choice of their most relevant publications. Applicants are ranked by an external committee, and the top candidates are invited to an interview and trial lecture, which are supervised by an internal committee. The internal committee conducts the final ranking (Frölich et al., 2018).

Each CV in the experiment included information on a number of attributes, most importantly the applicant's name (gender), whether the applicant has children (two or none), and the applicant's publication list. The respondents were randomized into four groups, A–D, and evaluated two CVs each. Group A evaluated a hypothetical male candidate without children, Group B a female candidate without children, Group C a male applicant with two children, and, finally, Group D a female applicant with two children. The two CVs were identical along the dimensions gender and number of children, but, as mentioned above, we varied the quality of the CV. The first CV had four publications, while the second had six publications. All CVs had at least two articles published in well-known journals in the respective fields. The randomization of gender and number of children means that we can make causal interpretations of the effect of these variables on the evaluation of the CVs.

The effects of gender and children were evaluated using a between-subjects design, while the quality of CV effects was evaluated using a within-subjects design. The between-subjects part of the design has an important advantage, as it helps conceal that the main intent of the project is to evaluate the between-subject treatments (gender and children). If the respondents would understand the purpose of the experiment, they may give socially desirable answers and create a bias in our

estimates. Moreover, the between-subjects part of the design makes the data we collect less sensitive because we will never measure a gender bias of a particular subject. For a particular subject, the data only reveal the subject's ranking of the publication list. Effects of the quality of the CV were evaluated using a within-subjects design.

Our research design has several advantages over other designs. With observational studies, it has proven to be very difficult to measure unequal treatment in a credible manner due to confounding factors. With laboratory experiments, it is difficult to create realistic situations to study, where the results also apply outside of the laboratory. With interviews with job applicants, it is difficult to know if the answers reflect real or perceived discrimination and with interviews with faculty members, one cannot be sure that the answers are consistent with their behaviour. A survey experiment of gender bias addresses many of the problems associated with these other methods. Although we will not study real hiring processes, our research design will be useful to substantiate whether discrimination is likely to occur in real hiring processes at the sampled universities.

Before collecting the data, we described all aspects of the research design, variable operationalizations, model specifications, robustness checks, and handling of missing data in a preregistered research plan that was submitted to the Evidence in Governance and Politics' (EGAP) research register.

Choice of Subjects and Institutions

We started from three criteria as guidelines when choosing disciplines. The first was to include disciplines with a low share of female professors. The second criterion was to include disciplines with a high share of female PhD students relative to the share of female professors. These first two criteria indicate problems in recruiting female professors. The third criterion was that the number of researchers in the discipline, i.e., potential participants, had to be sufficiently large to substantially contribute to the data collection. In practice, it turned out that the third criterion was most important. The reason is that many broader disciplines consist of sub-fields in which very specific CVs have to be used, meaning that, if included, we would have had to treat them as separate disciplines and create a vast number of specific CVs for these sub-fields. Therefore, in the end, the chosen disciplines—Economics, Law, Physics, Political Science, Psychology, and Sociology—were those with many employed researchers, and they should give a fairly representative picture of the prevalence of gender bias in Iceland, Norway, and Sweden. Figure 1 shows the share of female professors in the included disciplines. We included all large institutions in the three countries in the survey experiment. Altogether, 17 institutions were included. We pool the data from the three countries, since the sample size is too small to analyse the data by country.

Construction of Hypothetical CVs

To create realistic CVs, we studied a large number of real CVs in each discipline available online, typically at researchers' personal homepages. We also consulted colleagues in each discipline to give feedback on the CVs. Finally, in March 2019, we conducted a pilot study ($N = 22$) at a Swedish university not part of the main study. The first aim of the pilot was to check that the distribution of the outcome variables was reasonable centred on the answering scale, 1–7. We found that the mean of the competence and hireability indices were 4.5 and 4.0, respectively. The second aim was to gather comments and feedback through an open question at the end of the survey. After the pilot study, we made several minor changes of the survey questions and the CVs.

Each constructed CV fits on a single page to facilitate that the participants easily can overview the content. At the very top of the CVs, there is personal information such as the candidate's name, birth date, nationality, and civil status (marital status and the number of children). Next follows information about the candidate's research interests, employment history, education, teaching experience, experience of professional services and memberships in research networks and organizations, and, finally, publications in international peer-reviewed journals.

The experimental manipulations are the name of the applicant, which signals gender, civil status, which states that the candidates are married and have either no children or two children, and the publication list, which contains either four or six international peer-reviewed publications. Supplementary Figures S1 and S2 in Appendix E show two example CVs.

Outcomes and Other Variables

We had two main outcomes in the experiment. The first was an evaluation of the competence of the applicant. Our measure of competence is an additive index constructed from the answers to the following three questions (see the preregistered research plan): (i) *To what extent do you consider the applicant as competent for the position?*, (ii) *To what extent do you consider the applicant to possess the necessary qualifications for the position?*, and (iii) *How qualified do you consider the*

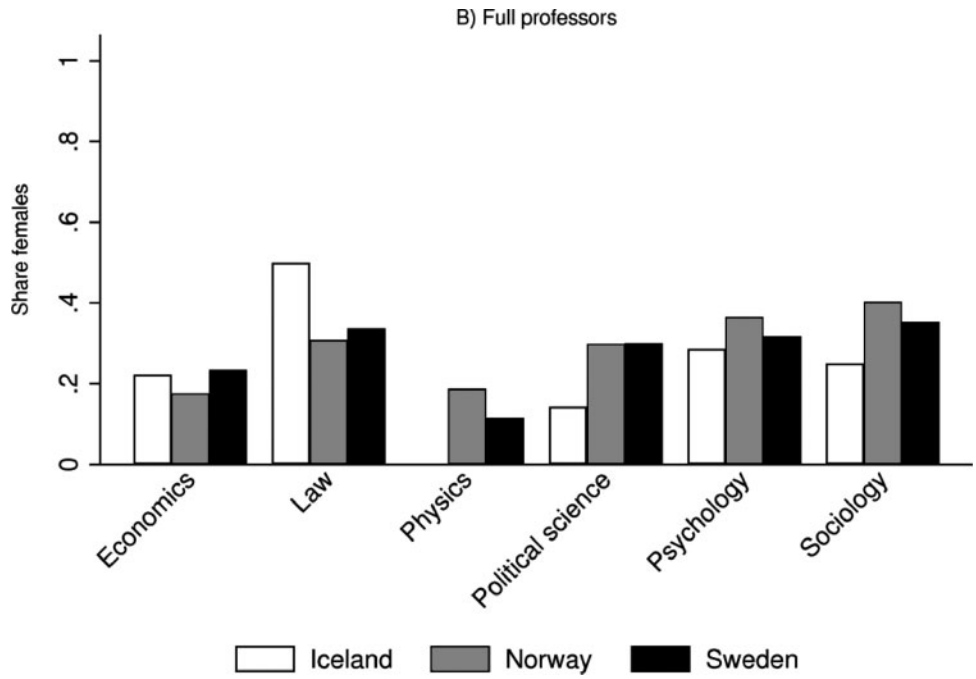

Figure 1. Share of women in Full Professor positions in Iceland, Norway, and Sweden, by discipline, 2017
Notes: In Iceland and Sweden, Economics include both Economics and Business Administration. In Norway, the numbers include both Full Professors and Docents, the latter being an academic appointment between Associate Professor and Full Professor. Data for Iceland are based on our own compilations of staff lists available at the webpages of all Icelandic universities. Data for Sweden are obtained from the Swedish Higher Education Authority. Data for Norway are obtained from the Nordic Institute for Studies in Innovation, Research and Education (NIFU).

applicant to be? The respondent answered the questions on a scale 1–7 where 1 is the lowest score and 7 is the highest score. The second outcome was an evaluation of the hireability of the applicant. Our measure of hirability is an additive index constructed of the answers to following three questions: (i) *How likely is it that you would hire the applicant?*, (ii) *How likely is it that you would invite the applicant to a job interview?*, and (iii) *What is your assessment of the probability that the applicant will get the job?* The respondent answered these questions too on a scale 1–7. The distributions of the outcome variables competence and hireability are shown in Supplementary Figure S5 in Appendix F.

We also asked the respondents questions about background characteristics. These were: gender (a binary indicator for being female or male), age (measured as a continuous variable), number of years since obtaining the PhD degree (a continuous variable), number of years employed at the current institution (a continuous variable), whether having participated in a hiring committee for permanent positions during the last 5 years (a binary variable; yes or no), faculty position (a binary indicator for Full Professor or not) and citizenship (a binary indicator for being a citizen in the survey country, e.g., being an Icelandic citizen and participating in the survey in Iceland). These variables are used for balance and robustness checks (see Supplementary Appendix B).

Sampling

We employed two approaches to collect data to ensure that enough respondents participated in the study. In the first approach, research assistants collected emails for all relevant faculties at the largest universities in Iceland, Norway, and Sweden. The participants were then recruited by email to fill out an electronic survey in the survey software Qualtrics (www.qualtrics.com). In the second approach, the participants were recruited at internal faculty meetings. We contacted departments in the universities and asked for access to internal meetings where we could distribute the survey. Respondents were asked to fill out the survey, using pen and paper, on the spot without communicating with others. Out of the 775 participants, 706 were recruited through the first method and 69 through the second method.

Altogether, the targeted population consists of approximately 2,000 individuals, of whom 775 (39 per cent) participated in the study and voluntarily ranked applications. The number of responses varied from 93 in economics to 155 in psychology. Among those who answered the background questions in the survey, 66 per cent are male, the average age is 49, 59 per cent have experience from hiring committees, and 46 per cent are full professors. Eight per cent, 43 per cent, and 51 per cent work at an Icelandic, Norwegian, and Swedish university, respectively. Respondents self-select to participate in the survey, but the rather high response rate and the descriptive statistics suggest that the sample is fairly representative of the population. Also, since we conducted an experiment, self-selection to the study will not harm the internal validity, but might limit the external validity, of the results.

Data Analysis

We estimated the treatment effects in a regression framework. To test H1, we ran the following regression separately both on evaluations of the baseline CV with four publications ($N = 775$) and on evaluations of all CVs ($N = 1{,}550$), i.e., the pooled sample of baseline CVs and CVs with a strong publication list:

$$y_{ijk} = \beta_F \text{FEMALE}_{jk} + \epsilon_{ijk} \quad (1)$$

In this model, i indexes respondent, j discipline, and k institution. The FEMALE dummy is equal to 1 for those receiving the gender treatment (CV of type B and D). Hypothesis H1 (that females have lower ratings than males) implies that $\beta_F < 0$.

To test H2, we ran the following regression using evaluations of the baseline CV and the pooled sample of all CVs:

$$\begin{aligned} y_{ijk} = {} & \pi_F \text{FEMALE}_{jk} + \pi_C \text{CHILDREN}_{jk} \\ & + \pi_{FC} \text{FEMALE}_{jk} * \text{CHILDREN}_{jk} + \epsilon_{ijk} \end{aligned} \quad (2)$$

In this model, CHILDREN is equal to 1 for groups receiving the children treatment (CVs of type C and D). Hypothesis H2 (that females have a lower return to children than males) implies that $\pi_{FC} < 0$.

To test H3, we run the following regression:

$$\begin{aligned} y_{ijk} = {} & \pi_F \text{FEMALE}_{jk} + \pi_C \text{CHILDREN}_{jk} + \pi_V \text{CV}_{ijk} \\ & + \pi_{FC} \text{FEMALE}_{jk} * \text{CHILDREN}_{jk} \\ & + \pi_{FV} \text{FEMALE}_{jk} * \text{CV}_{ijk} + \epsilon_{ijk} \end{aligned} \quad (3)$$

In this model, CV is equal to 1 for the evaluation of the strong CV and otherwise 0. Hypothesis H3 (that females have a lower return to a strong CV than males) implies that $\pi_{FV} < 0$.

Results

We analysed treatment effects using OLS regressions. Figure 2 displays the main results graphically (while formal tests of whether differences are statistically significant are presented in Table 1). The dots in the figure show the mean scores for male and female CVs, while the lines show the confidence intervals. Panel A in Figure 2 presents results for the competence ratings, using only the baseline CV with four publications (to the left) and using all CVs, i.e., with four and six publications (to the right). Panel B in Figure 2 shows the corresponding results for the hireability ratings. Contrary to our expectations, we find that, for both competence and hireability, female CVs get higher ratings than male CVs. However, as expected, the average ratings for both males and females are higher when we include the CVs with six publications (to the right in the figure).

Table 1 shows the formal tests of whether female CVs are evaluated differently from male CVs. Again, panel A Table 1 uses only the baseline CVs, while panel B uses the pooled sample of CVs. In these regressions, the constants are the average evaluations of the male candidate, while the coefficient for a female CV shows the difference in rating for the female candidate compared to the male candidate. In panel A, we see that, on average, a male candidate's CV is rated 4.26 on competence and 3.77 on hireability and a female candidate's CV is rated 0.29 higher on competence and 0.28 higher on hireability. Both differences are statistically significant (at the 1 and 5 per cent level, respectively). In panel B, using all observations, we obtain very similar results, but the precision is slightly better.

In Supplementary Appendix B, we present the results of the preregistered balance and robustness tests. They show balance across treatment groups for all background variables except gender, which most likely is the result of an unlucky draw from the population. In practice, this imbalance for gender seems unimportant, since the results remain unchanged when we control for gender and other background characteristics of the participants (for details, see Balance and Robustness Checks section in Supplementary Appendix B). Moreover, we repeat the regressions in Table 1 including a dummy for whether the survey was answered by paper and pencil or online, discipline fixed effects, institution fixed effects, and discipline times institution fixed effects. These variables could increase the precision of the estimates and change the estimated treatment effects if there is an imbalance in the share of treated in different groups of respondents (e.g., age or gender groups). The treatment effect barely moves across the robustness checks. We therefore reject H1.

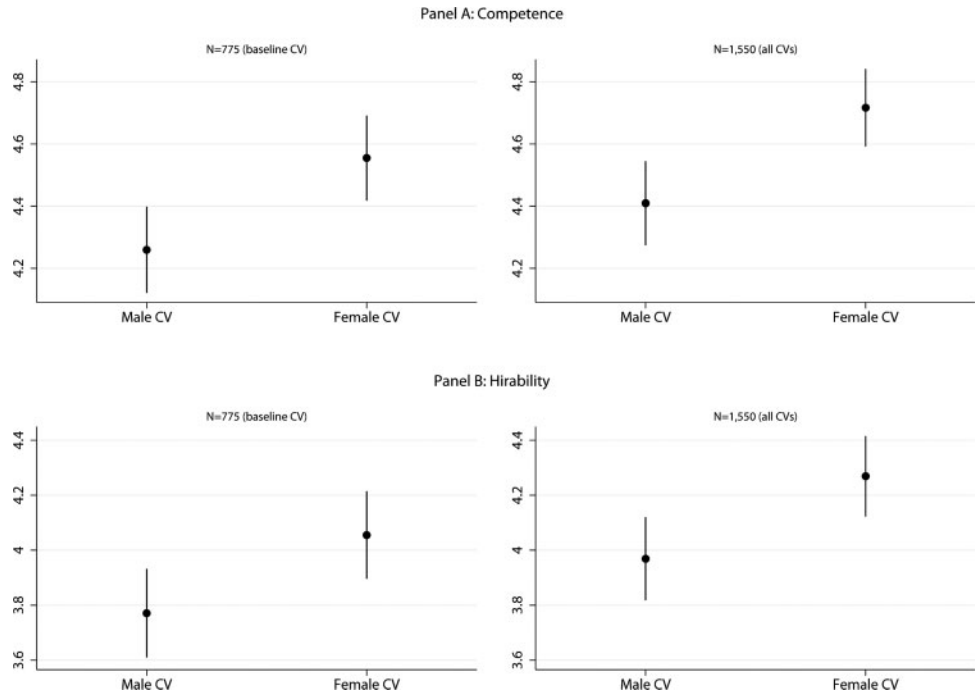

Figure 2. Ratings of CVs for male and female candidates (Panel A: Competence. Panel B: Hirability).
Note: The gender difference in each graph is statistically significant (see Table 1).

Table 1. Ratings of CVs for male and female candidates

	Competence (1)	Hireability (2)
Panel A ($N = 775$):		
Female CV	0.2949***	0.2842**
	(0.0992)	(0.1153)
Constant (male CV)	4.2596***	3.7710***
	(0.0707)	(0.0821)
Panel B ($N = 1{,}550$):		
Female CV	0.3075***	0.2999***
	(0.0937)	(0.1070)
Constant (male CV)	4.4095***	3.9689***
	(0.0688)	(0.0769)

Notes: The regressions include no other covariates than the female CV dummy. In Panel B, standard errors are clustered by respondent.
***Significant at the 1 per cent level,
**significant at the 5 per cent level;

If we estimate the treatment effect for H1 separately by academic field, we find that the treatment effect estimates are smaller for Law and higher for Physics and Psychology. However, a formal test does not rule out that the treatment effect is the same across fields (see Supplementary Table S3 in Appendix A).

Finally, we find no support for H2 and H3. The treatment effect estimates are always small (below 0.1) and statistically insignificant (see Supplementary Tables S1 and S2 in Appendix A). Thus, somewhat surprisingly, we find no evidence that the pay-off for children and for a strong CV is lower for women than for men. This conclusion remains across the robustness and specification checks.

Discussion and Conclusion

For decades, scholars have debated whether bias against women explains the persistent gender gap in top-level academic positions. This study has examined the role of bias in the critical moment when candidates are evaluated for positions as Associate Professors. In the Nordic countries, landing a position as an Associate Professor translates into a permanent career in academia and is a necessary stepping-stone for later promotion to the position as Full Professor. To the extent that female researchers experience bias in this crucial stage in their academic careers, it would help explaining why women are underrepresented at the top of the academic hierarchy even in the 'women-friendly' Nordic region (Borchorst and Siim, 2002; Nielsen, 2017).

Despite mixed results in the existing experimental literature on gender bias in academic recruitment (Moss-

Racusin *et al.*, 2012; Williams and Ceci, 2015; Ceci, 2018), we expected our survey experiment to reveal a male advantage in the evaluation of candidates to Associate Professor positions. We also expected female candidates to have a lower return to children and strong CVs than males. The main rationale behind these hypotheses was the persistent gender gap at the top level in all countries and almost all fields examined, combined with theories of discrimination and stereotype content, which suggest that women—and especially women with children—would be viewed as less competent and hireable than men.

Contrary to our main hypothesis, however, we did not find evidence of a bias against female applicants to Associate Professor positions in the Nordic region. Rather, female candidates are perceived as both more competent and hireable compared to equally qualified male candidates. Furthermore, we find no evidence of a child penalty for neither male nor female applicants and no gender difference in the pay-off from a strong CV.

A potential limitation of the study is that we cannot rule out self-selection to the survey experiment. This might lower the external validity of the results as we cannot be certain that the participants' ratings fully reflect the attitude of the full population in the chosen disciplines. However, the rather high response rate (39 per cent) should make this less of a concern. Another limitation is that although we aimed for obtaining a fairly representative picture of the prevalence of gender bias in Iceland, Norway, and Sweden by including six disciplines with many employed researchers (Economics, Law, Physics, Political Science, Psychology, and Sociology), we cannot rule out the possibility that the inclusion of other disciplines would affect the results.

Our results stand in contrast to many previous studies of various gender disparities in academia, which suggest that bias against women is widespread (e.g., Wennerås and Wold, 1997; Barbezat and Hughes, 2005; Moss-Racusin *et al.*, 2012). However, our findings are nonetheless in line with a number of existing studies on academic hiring that suggest a female advantage (Baldi, 1995; Merritt and Reskin, 1997; Wolfinger, Mason and Goulden, 2008; National Research Council, 2009; Lutter and Schröder, 2016). In particular, our findings are consistent with Williams and Ceci (2015), the only existing experimental study that assesses the prevalence of gender bias in recruitment to tenure track positions. Yet, the female advantage we find is far less pronounced than the 2-to-1 female advantage that Williams and Ceci (2015) report. The smaller effects in our study could reflect differences in research design. Specifically, in Williams and Ceci (2015), the same subject evaluated both male and female applicants (while we used a between-subject design) and their subjects were asked to rate detailed narratives of exceptionally qualified applicants (instead of CVs). These differences could have led to a particularly strong assessment of the female candidates. Another important difference is that Williams and Ceci (2015) examined only STEM fields. Finally, differences in results could reflect that the studies were conducted at different times and in different institutional contexts.

What could explain our finding that, if anything, the bias is in *favour* of female candidates? This is a difficult question to answer fully. However, one possibility is related to the fact that the present study has been conducted in the Nordic region, well known for its institutionalized gender equality norms (Teigen and Skjeie, 2017). These norms could impact, e.g., hiring committees, making them aware of the demands in the Nordic region to hire more female academics. To the extent that this demand is widely accepted in academic communities, it is possible that the respondents have internalized this and consequently given more favourable evaluations of the female candidates.

Regardless of how we explain the results of this study, one important question remains: How can we account for the persistent gender gap in top-level academic positions if bias in recruitment is not the reason? One potential explanation is that the underrepresentation of women in professor positions is the result of sorting mechanisms occurring at earlier stages in the academic career, which—constrained or not—could lead fewer women to ascend to professor positions. Another potential explanation is that bias against female academics occurs earlier in career trajectory: The lack of bias against female applicants in our study does not rule out the possibility that men experience advantages in other phases of academic life, such as in monitoring, review boards, or peer-review assessments. Indeed, the virtue of experiments such as this is the ability to examine directly whether gender bias exists in evaluations of candidates of different gender, all else being equal. Bias in processes prior to the event of applying for a position as an Associate Professor is not studied in experiments of our kind, nor are sorting mechanisms that may lead more men than women to embark on an academic career. If such biases and sorting mechanisms are widespread, they might be an important explanation of the existing gender gap in academic top positions. Yet when female applicants have succeeded in getting on par with their male peers, our study suggests that women are not less likely than men to being awarded permanent positions in academia.

Supplementary Data

Supplementary data are available at *ESR* online.

Funding

This work was supported by NordForsk (grant number 80713) and is part of the Nordic Centre for Research on Gender Equality in Research and Innovation (NORDICORE).

Acknowledgements

We thank the two anonymous reviewers and the journal's editors for constructive comments to a previous version of the manuscript, and participants at the NORDICORE meetings at the Institute for Social Research in Oslo, Norway, for their feedback to the research design and analyses. Many thanks also to Ástrós Anna, Olof Axman, Live Kjos Fjell, Øyvind Skorge, Maya Staub, and Sahra Ali Abdullahi Torjussen for invaluable contributions in different stages of the research process.

References

Abele, A. E. *et al.* (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, **387**, 1063–1065.

Allen, M. and Castleman, T. (2001). Fighting the pipeline fallacy. In Brooks, A. and Mackinnon, A. (Eds.), *Gender and the Restructured University Changing Management and Culture in Higher Education*. Buckingham: The Society for Research into Higher Education & Open Press University. pp. 151–165.

Anders, S. (2004). Why the academic pipeline leaks: fewer men than women perceive barriers to becoming professors. *Sex Roles*, **51**, 511–521.

Arrow, K. J. (1973). The theory of discrimination. In Ashenfelter, O. and Rees, A.(Eds.), *Discrimination in Labor Markets*. Princeton, NJ: Princeton University Press, pp. 3–42.

Baldi, S. (1995). Prestige determinants of first academic job for new sociology Ph.D.s. *The Sociological Quarterly*, **36**, 777–789.

Barbezat, D. A. and Hughes, J. W. (2005). Salary structure effects and the gender pay gap in academia. *Research in Higher Education*, **46**, 621–640.

Bergman,S., & Rustad, L. M. (2013). The Nordic region – a step closer to gender balance in research? Joint Nordic strategies and measures to promote gender balance among researchers in academia. TemaNord Report 2013:544. Copenhagen: Nordic Council of Ministers.

Blau, F. D. *et al.* (2010). Can mentoring help female assistant professors? Interim results from a randomized trial. *American Economic Review*, **100**, 348–352.

Borchorst, A. and Siim, B. (2002). The women-friendly welfare states revisited. *NORA - Nordic Journal of Feminist and Gender Research*, **10**, 90–98.

Burgess, D. and Borgida, E. (1999). Who women are, who women should be: descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, **53**, 665–692.

Cech, E. A. and Blair-Loy, M. (2019). The changing career trajectories of new parents in STEM. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 4182–4187.

Ceci, S. J. (2018). Women in academic science: experimental findings from hiring studies. *Educational Psychologist*, **53**, 22–41.

Ceci, S. J. and Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 3157–3162.

Correll, S. J. (2004). Constraints into preferences: gender, status, and emerging career aspirations. *American Sociological Review*, **69**, 93–113.

Correll, S. J., Benard, S. and Paik, I. (2007). Getting a job: is there a motherhood penalty? *American Journal of Sociology*, **112**, 1297–1339.

Correll, S. J. and Ridgeway, C. L. (2003). Expectation states theory. In Delamater, J. (Ed.), *Handbook of Social Psychology*. New York, NY: Kluwer Academic/Plenum Publishers, pp. 29–51.

Cuddy, A. J. C., Fiske, S. T. and Glick, P. (2004). When Professionals Become Mothers, Warmth Doesn't Cut the Ice. Journal of Social issues, **60**, 701-718.

De Groot, J. (1997). After the ivory tower: gender, commodification and the 'academic'. *Feminist Review*, **55**, 130–142.

European Commission. (2019). *She Figures 2018*. Luxembourg: Publications Office of the European Union.

Fiske, S. T., Cuddy, A. J. and Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, **112**, 77–83.

Fiske, S. T. and Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 13593–13597.

Fiske, S. T. *et al.* (1991). Social science research on trial: use of sex stereotyping research in Price Waterhouse v. Hopkins. *American Psychologist*, **46**, 1049.

Fiske, S. T. *et al.* (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, **82**, 878.

Foschi, M., Lai, L. and Sigerson, K. (1994). Gender and double standards in the assessments of job candidates. *Social Psychology Quarterly*, **57**, 326–339.

Frölich, N. *et al.* (2018). *Academic Career Structures in Europe: Perspectives from Norway, Denmark, Sweden, Finland, the Netherlands, Austria and the UK*. Oslo: Nordic Institute for Studies in Innovation, Research and Education (NIFU).

Heckman, J. J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, **12**, 101–116.

Judd, C. M. *et al.* (2005). Fundamental dimensions of social judgment: understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, **896**, 899.

Lutter, M. and Schröder, M. (2016). Who becomes a tenured professor, and why? Panel data evidence from German sociology, 1980–2013. *Research Policy*, **45**, 999–1013.

Lutter, M. and Schröder, M. (2019). Is there a motherhood penalty in academia? The gendered effect of children on academic publications in German sociology. *European Sociological Review*, **36**, 442–459.

Merritt, D. J. and Reskin, B. F. (1997). Sex, race, and credentials: the truth about affirmative action in law faculty hiring. *Columbia Law Review*, **97**, 199–311.

Miller, J. and Chamberlin, M. (2000). Women are teachers, men are professors: a study of student perceptions. *Teaching Sociology,* **28**, 283–298.

Monroe, K. *et al.* (2008). Gender equality in academia: bad news from the trenches, and some possible solutions. *Perspectives on Politics*, **62**, 215–233.

Moss-Racusin, C. *et al.* (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 16474–16479.

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

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, **56**, 799–866.

Nielsen, M. W. (2017). Scandinavian Approaches to Gender Equality in Academia: A Comparative Study. Scandinavian Journal Ofeducational Research, **61**, 295-318.

OECD. (2018). *Education at a Glance 2018*. Paris: OECD Publishing.

Orupabo, J. (2018). Cultural stereotypes and professional self-socialisation in the transition from education to work. *Journal of Education and Work*, **31**, 234-246.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, **624**, 659–661.

Reskin, B. F. and Roos, P. A. (1990). *Job Queues, Gender Queues: Explaining Women's Inroads into Male Occupations*. Philadelphia, PA: Temple University Press.

Rich, J. 2014. What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000. IZA DP No. 8584.

Ridgeway, C. and England, P. 2007. Sociological approaches to sex discrimination in employment. In *Sex Discrimination in the Workplace*. Oxford, Blackwell.

Ridgeway, C. L. (2011). *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford: Oxford University Press.

Ridgeway, C. L. and Bourg, C. (2004). Gender as status: an expectation states theory approach. In Eagly, A. H., Beall, A. E. and Sternberg, R. J. (Eds.), *The Psychology of Gender*. New York, NY: The Guilford Press, pp. 217–241.

Rivera, L. A. (2017). When two bodies are (not) a problem: gender and relationship status discrimination in academic hiring. *American Sociological Review*, **82**, 1111–1138.

Santos, G. and Dang Van Phu, S. (2019). Gender and academic rank in the UK. *Sustainability*, **11**, 3171.

Steinpreis, R., Anders, R. K. and Ritzke, K. D. (1999). The impact of gender on the review of the CVs of job applicants and tenure candidates: a national empirical study. *Sex Roles*, **41**, 509–528.

Teigen, M., and Skjeie, H. (2017). The Nordic gender equality model. In Knutsen, O. (Ed.), *The Nordic Models in Political Science. Challenged but Still Viable*? Oslo: Fagbokforlaget. pp. 125–147.

van Hek, M., Kraaykamp, G. and Wolbers, M. H. J. (2016). Comparing the gender gap in educational attainment: the impact of emancipatory contexts in 33 cohorts across 33 countries. *Educational Research and Evaluation*, **22**, 260–282.

Wennerås, C. and Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, **387**, 341–343.

Williams, W. M. and Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 5360–5365.

Wolfinger, N. H., Mason, M. A. and Goulden, M. (2008). Problems in the pipeline: gender, marriage, and fertility in the ivory tower. *Journal of Higher Education*, **79**, 388–405.

World Economic Forum. (2020). *The Global Gender Gap Report 2019*. Geneva: World Economic Forum, available from: http://repor ts.weforum.org/global-gender-gap-report-2020/ [accessed 6 April 2020].

Xu, Y. J. (2008). Gender disparity in STEM disciplines: a study of faculty attrition and turnover intentions. *Research in Higher Education*, **49**, 607–624.

Magnus Carlsson is an associate professor in economics at Linnaeus University, Sweden. Current research interests include labour economics and applied microeconomics, and a large part of his research aims at studying discrimination and which factors are important in explaining observed group differences in labour market outcomes. Carlsson's work has been published in journals such as *The Review of Economics and Statistics*, *Labour Economics*, and *ILR Review*.

Henning Finseraas is an associate professor in political science at the Norwegian University of Science and Technology (NTNU) and a research professor at the Institute for Social Research in Oslo, Norway. Current research interests include electoral behaviour and public opinion, and quantitative evaluations of public policy, often related to immigration, political economics, and the welfare state. Finseraas' work has been published in the journals *American Journal of Political Science*, *British Journal of Political Science*, and *Journal of Public Economics*.

Arnfinn H. Midtbøen is an associate professor in sociology at the University of Oslo and a research professor at the Institute for Social Research in Oslo, Norway. Current research interests include economic and political sociology, and frequent topics are labour market discrimination, citizenship policies, and incorporation patterns among descendants of migrants. Midtbøen's work has been published in journals such as *Proceedings of the National Academy of the Sciences (PNAS)*, *British Journal of Sociology*, and *International Migration Review*.

Guðbjörg Linda Rafnsdóttir is a pro-rector of science and a professor in sociology at the University of Iceland. Current research interests include gender segregation, gender quotas, and work/family balance. Rafnsdóttir's work has been published in journals such as *Gender Work and Organization*, *Politics and Gender*, and *Work, Employment and Society*.