# Gender bias in evaluating assistant professorship applicants? Evidence from harmonized survey experiments in Germany and Italy

Klarita Gërxhani [a,b,1], Nevena Kulic [c,1,*], Alessandra Rusconi [d,1], Heike Solga [d,e,1]

[a] *School of Business and Economics, Free University Amsterdam, the Netherlands*
[b] *Tinbergen Institute, the Netherlands*
[c] *University of Pavia, Italy*
[d] *WZB – Berlin Social Science Center, Germany*
[e] *Freie Universitaet Berlin, Germany*

A B S T R A C T

This study investigates gender biases in the evaluation of applicants for assistant professorships in Germany and Italy. Drawing on the justification-suppression model of prejudice expression, we explore whether biases against women are expressed, suppressed, or even reversed in the appointment process, considering the different normative gender climates and gender equality strategies in the two countries. Using harmonized factorial survey experiments with professors of economics, political science, and social sciences, we found that women in Germany have an advantage both in perceived qualification for an assistant professorship and in the propensity to receive an interview invitation. In contrast, women in Italy are neither disadvantaged nor advantaged. We also examine whether gender biases exist when there is ambiguity about applicants' academic performance (co-authorship) and career commitment (parental leave). Our results reveal a co-authorship penalty and a parenthood premium in both countries, with no gender differences observed. Our exploratory country comparison suggests that Germany's pro-active gender equality policies may be more effective in reducing the gender gap in assistant professor appointments compared to Italy's gender-neutral approach, by favoring equally qualified female applicants.

## 1. Introduction

Gender inequalities in academia, especially the underrepresentation of women among professors, have attracted significant attention in recent decades (European Commission, 2020; LERU, 2018). While there is a broad consensus on the need to increase the proportion of female professors, an ongoing debate persists regarding the reasons for their underrepresentation. Are women less likely to apply for professorships, or are they less likely to be hired when they do apply? The prevailing view suggests that biases and discrimination in the recruitment process are major contributors to women's underrepresentation among professors (LERU, 2018; NASEM, 2020).

---

* Corresponding author. Università di Pavia, Dipartimento di Scienze Politiche e Sociali, C.so Strada Nuova 65, 27100, Pavia, Italy.
*E-mail address:* nevena.kulic@unipv.it (N. Kulic).
[1] The order of the authors is alphabetical. All authors contributed equally to this article.

This study contributes to this debate by examining whether gatekeepers (i.e., university professors as members of appointment committees) in Italy and Germany evaluate female and male applicants for assistant professorships differently. By assessing the presence of gender bias in assistant professor appointments in the two countries, we explore how differences in institutional and normative contexts influence the evaluation of female and male applicants. We conducted harmonized factorial survey experiments and asked professors across three social sciences disciplines (economics, political science, and the social sciences) to rate the academic profiles of hypothetical applicants that varied by gender and signals of productivity and commitment. These evaluations focused on two dimensions: perceived qualifications for the assistant professor position and the likelihood of inviting the applicant for a job interview. Both dimensions are essential for hiring and promotion practices, as the former relates to beliefs about an applicant's competence, while the latter reflects the second step of the selection process. We expected institutional and normative contexts to have distinct impacts on both outcomes in each country.

Studies on gender bias in academia have predominantly focused on tenured (associate and full) professor appointments (see Table A1 in the Appendix for a summary of key information on these studies). While some found discrimination against women (e.g., in the US, see Ginther and Kahn, 2021; Perna, 2001; Sarsons, 2017; Weisshaar, 2017; in Europe, see Filandri and Pasqua, 2021; Ooms et al., 2019), others observed the opposite (see Ceci et al., 2023). Recent studies focusing on Europe have either found no gender differences (e.g., Aksnes et al., 2022; Bosquet et al., 2019; Ginther and Kahn, 2021; Moratti, 2021) or reported a female advantage for tenured professorships (e.g., Auspurg et al., 2017; Carlsson et al., 2021; Henningsen et al., 2022; Lutter et al., 2022; Solga et al., 2023).

However, these studies may not capture the whole picture, as they do not account for the possibility that more women than men may exit academia due to discriminatory recruitment practices at earlier stages in their careers. The transition to assistant professorships is crucial for understanding women's underrepresentation (Williams and Ceci, 2015), as assistant professors serve as a vital pool for future associate and full professors. Unlike lower-level hiring for post-docs and research assistants, assistant professor recruitment typically involves more formalized procedures (Herschberg, Benschop, and van den Brink, 2018). Standardized and transparent processes are expected to ensure merit-based decision-making and reduce gender bias. Despite the importance of this career stage, very few studies have addressed it, with inconsistent findings (see Appendix, Table A1). For instance, an older experimental study in the US (Steinpreis et al., 1999) and a more recent observational study with pooled data from one German and one Dutch technical university (Ooms et al., 2019) found evidence of discrimination against women. Additionally, Gërxhani et al. (2023) examined gender bias in evaluations of collaborative work in Italy through a vignette experiment involving professors working in the humanities and social sciences and found double standards against female applicants when they were evaluated by male professors, particularly in the field of economics. In contrast, several recent experimental studies reported a female advantage in evaluation ratings in the US (Carey et al., 2020; Ceci and Williams, 2015; Williams and Ceci, 2015). These contradictory findings regarding discrimination against women in academic hiring may be explained by differences in the focuses of the analyses, in the countries, disciplines, and time periods studied.[2]

We contribute to this limited body of research by investigating whether gender bias varies depending on a country's normative and institutional context. We expand on the most recent study on Italy by focusing specifically on the social sciences and considering the role of gender alone in the hiring process, along with signals of career commitment and research productivity. Most importantly, we expanded the geographical focus to include two countries. The few studies that have included multiple countries primarily relied on pooled analyses, which hide country-specific differences in gender policy approaches. Additionally, comparing results from single-country studies is challenging due to issues with harmonization. In this study, we focused on Germany and Italy, two European countries that exhibit differences in gender policies and norms. Whereas Germany has implemented a wide range of gender-equality policies, including *gender-preferential selection* in academic hiring, which requires favoring women when applicants are equally qualified, Italy has adopted gender-neutral measures that emphasize standardized eligibility requirements in academia to mitigate biased interpretations of productivity. Although both countries generally exhibit traditional gender norms, Germany tends to have more progressive gender attitudes than Italy.

We define gender bias in evaluations as the differences in assessments of male and female applicants when all other factors are equal. This bias stems from the gender stereotypes and prejudices held by evaluators. The prevailing view on discrimination against women posits that these stereotypes are pervasive and expressed in evaluative judgments. However, Crandall and Eshleman's (2003) justification-suppression model of prejudice expression offers an alternative perspective. While gender stereotypes may indeed be prevalent, they may not be expressed in evaluations, resulting in no observable gender bias. This suppression can occur due to internal motivations driven by a commitment to gender equality or can be externally motivated because of social control or a gender-egalitarian normative climate in which, for example, expressing gender stereotypes is considered socially unacceptable.

It is reasonable to expect that a gender-egalitarian normative climate is more prevalent today, partially due to the implementation of national affirmative action policies by governments and universities. These policies aim to increase the representation of women in academia and prevent unconscious biases from emerging in recruitment practices (see Crowley and Sansonetti, 2019; LERU, 2018). Examples of such policies include gender quotas on selection committees, gender-sensitive recruitment protocols, and the appointment of equal opportunity officers (European Commission, 2014). In such contexts, one may even observe instances of bias reversal, where female applicants receive higher ratings in evaluations. This is not necessarily because male applicants are perceived as less competent but rather because female applicants are seen as bringing additional "diversity value that goes above and beyond their human capital

---

[2] Indeed, a recent cross-temporal meta-analysis of gender bias in hiring decisions across different countries and labor market segments observed reductions in bias against women and even occasional reversals of bias against men in gender-mixed and male-stereotypical jobs in the last decade, potentially influenced by ideologies like feminism and egalitarianism (Schaerer et al., 2023).

value" to the organization (Schearer et al., 2023, p. 3; see also Chang et al., 2019; Leslie et al., 2017). Furthermore, a preference for gender equality in outcomes favors equally qualified minority-gender applicants (Farré and Ortega, 2023), thereby benefiting female applicants to address the historical underrepresentation of women in academia. The pursuit of these goals would also involve assigning higher ratings to female applicants in evaluation processes.

We employed factorial survey experiments to study gender bias in professorial appointments due to several methodological advantages. First, because there are real-world differences in how individuals of different genders apply for positions (Urry, 2015), such experiments standardize academic profiles and application behaviors for both male and female applicants. Consequently, any observed differences in the evaluations can be attributed solely to gender bias. Second, harmonized experiments create identical applicant pools in the two countries. In 2017, approximately 46% of assistant professors in the social sciences were women in Germany, compared to 49% in Italy (Statistisches Bundesamt, 2018, p. 48, 52; Morana and Sagramore, 2020, p. 11).[3] However, these figures say little about the prevalence of gender biases in recruitment decisions, as they reflect the combined results of application and recruitment behaviors. Indeed, the pool of potentially eligible women differs between the two countries: In the PhD graduation cohorts of 2012 and 2016, which contributed to assistant professor hiring by 2017 (with fast or slower transitions), the proportion of female graduates in the social sciences was lower in Germany (42% and 54%, respectively) than in Italy (53% and 59%, respectively) (European Commission, 2016, 2019). Third, factorial survey experiments align well with the justification-suppression model mentioned above, as they allow us to observe actual rating behaviors and, thus, the expression of potentially existing prejudices rather than the presence of prejudices per se.

## 2. The German and Italian context

We provide some information on the German and Italian context. Our focus on the academic labor market increases comparability. In both countries, openings for professorships must be advertised publicly and internationally. Departments form appointment committees that review applications, select candidates for interviews, request external reviews, and make the final rankings. The key criteria for productivity, such as publications in high-ranked international journals and presentations at international conferences, are similar across both countries. Moreover, both academic systems share two characteristics that aid their comparability. First, all assistant professors in Germany (since 2002) and Italy (since 2010) are hired on temporary contracts—with or without tenure track. Second, the number of associate and full professorships is limited in both countries; thus, access to tenured professorships is only possible if an "empty" slot is available.

The two countries differ, however, in their normative gender climates. For example, our descriptive analysis of respondents to the European Values Study (EVS) 2017 with tertiary degrees found that only 19% of Germans (strongly) agreed with the statement, "When a mother works for pay, the children suffer," compared to 31% of Italians. Additionally, fewer Germans (32%) (strongly) agreed that "All in all, family life suffers when the woman has a full-time job," while 44% of Italians held this view. Furthermore, fewer Italians (73%) (strongly) disagreed with the statement, "When jobs are scarce, men have more rights to a job than women," compared to Germans (90%).[4] Moreover, the 2018 Global Gender Gap Index, which measures progress toward gender equality, reported a lower score for Italy than for Germany (0.71 versus 0.79) (Kosakowska-Berezecka et al., 2023). Italy also scored higher than Germany on various traditional gender ideology indicators, such as subjective beliefs regarding power distributions (0.5 versus 0.35) (Kosakowska-Berezecka et al., 2023), and on precarious masculinity beliefs, indicating the glorification of masculinity (4.24 versus 3.41) (Bosson et al., 2021).

Regarding approaches to gender equality in higher education institutions, 89% of German higher education institutions have implemented gender equality measures compared to only 56% of Italian institutions (European Commission, 2021, p. 170).[5] A closer look at specific policies concerning professorial appointments also reveals clear differences. Germany has a law (first implemented in 1989) that requires organizations to favor women when men and women perform equally, provided women are underrepresented in the area in question. This is known as "weak gender-preferential selection," as opposed to "strong preferential selection," which applies to female applicants who are less qualified than their male counterparts (Silva et al., 2021). This law has been accompanied by other gender equality measures within universities, such as gender-sensitive appointment guidelines, the inclusion of equal opportunity officers on appointment committees to ensure compliance with gender equality standards, and implicit bias training for researchers, including professors. Since 1989, the Joint Science Conference (*Gemeinsame Wissenschaftskonferenz*) has regularly published a gender report to monitor progress. In 2008, the German Research Foundation (DFG), the country's most important research funding organization, introduced "Research-Oriented Gender Equality Standards" with the goal of increasing the representation of women among professors. As DFG members, all public universities have voluntarily committed themselves to these standards. Overall, the policies implemented at German universities can be understood as a *direct* strategy of gender-awareness management aimed at promoting

---

[3] Germany and Italy had similar shares of female full professors in 2018: 24% and 28%, respectively (European Commission, 2021, 190).

[4] EVS (2020). European Values Study 2017: Integrated Dataset (EVS, 2017). GESIS Data Archive, Cologne. ZA7500 Data file Version 4.0.0, https://doi.org/10.4232/1.13560.

[5] Before adopting an overall national strategy on gender mainstreaming in July 2021, Italy relied on regional or sector-specific good practices (see https://eige.europa.eu/gender-mainstreaming/countries/italy).More specifically, the National Research Programme for 2021–2027, as the primary government document for research and development (R&D), advises ensuring gender balance on recruitment and selection panels and encourages research institutions to promote gender equality in research without imposing any specific policies (https://eige.europa.eu/gender-mainstreaming/toolkits/gear/legislative-policy-backgrounds/italy?language_content_entity=en).

women in academic hiring (Henningsen et al., 2022).

In contrast, Italy has implemented gender-neutral policies, employing an *indirect* gender strategy that promotes "meritocratic" procedures intended to reduce biased interpretations in academic performance evaluations (Marini and Meschitti, 2018; van den Brink et al., 2010). Gender-neutral hiring policies have traditionally been favored in academia in general as a way to eliminate favoritism based on specific group membership. These policies, which gained popularity in the 1990s (Glazer, 2005), emphasize excellence and impose standardized criteria in hiring that equally apply to everyone independently of individual situations. However, while designed to be impartial, gender-neutral policies may not yield "neutral" outcomes and could instead result in institutional discrimination, favoring male-typical career paths by overlooking differences in men's and women's life courses (e.g., Orupabo and Mangset, 2022).

Thus, efforts to promote gender equality in Italy have been incorporated into wider efforts to eliminate all forms of clientelism within academia, leading to the introduction of the national scientific habilitation (*abilitazione scientifica nazionale,* or ASN) (Marini and Meschitti, 2018). A high degree of standardization in the overall habilitation criteria aims to promote "meritocracy" in professorial appointments based on quality (Murgia and Poggio, 2019). Since the introduction of the Gelmini law (210/2010) in 2012, the appointment of professors has entailed a two-stage procedure. First, the national habilitation process, accredited by the national ASN system, serves as a "fit-for-the-role filter" to select potential applicants for professorships (Marini and Meschitti, 2018, p. 898). Habilitation is based on a quantitative and qualitative assessment of the candidate's output, with a particular focus on the number and quality of publications and citation scores. Only those with a habilitation degree are eligible for the second stage, in which departments advertise positions and form independent committees that select from the pool of eligible applicants. These committees operate with greater autonomy and less transparency since habilitation is only an initial application "threshold," leaving the evaluation of candidates' scientific careers and career progression to the committees' discretion. Thus, while the first step is relatively standardized and aligns with calls "for more transparency and accountability in academic recruitment and selection" (van den Brink et al., 2010, p. 1459), thereby potentially reducing gender bias in evaluations of applicants' qualification, the second step leaves more room for gender bias to influence the recruitment process and, ultimately, the selection of appointees (Marini and Meschitti, 2018).

The distinct profiles of Germany and Italy, with their notable differences in traditional values and gender policies, lead us to investigate the moderating influence of normative and institutional contexts. We expect that these differences will affect our two key outcomes—perceived qualification and invitation propensity—in divergent ways. In Italy, where traditional gender norms are more ingrained, the gender-neutral "eligibility threshold" approach, which focuses only on defining and standardizing required qualifications, may help suppress (but not reverse) gender prejudice in the assessment of perceived qualifications; however, it may not influence invitation decisions since there is no requirement to favor women over men in hiring decisions, even among equally qualified applicants. Conversely, in the context of more progressive gender norms, Germany's direct approach to promoting gender equality through wide-ranging policies and measures may result in reversing gender biases in favor of women for both outcomes. Given Germany's "weak gender-preferential selection" law, which explicitly requires selecting a woman in the case of equally qualified candidates of opposite genders, the female advantage may be even greater for invitation propensity than for perceived qualification.

## 3. Theoretical considerations and hypotheses

### 3.1. The role of gender in evaluations

Based on the discussion in the previous section, our baseline assumption is that gender prejudices still, to some degree, persist in both Germany and Italy. Consistent with several of the studies mentioned above highlighting the discrimination of women in academia, statistical discrimination theory would predict biases against women in assistant professor recruitment because evaluators may consider women's average productivity to be lower than men's (e.g., because of caregiving responsibilities) (Arrow, 1973; Phelps, 1972). Thus, evaluators may view women as less qualified and hire them less frequently. Even if evaluators have "perfect" or full information about women's productivity, status characteristics theory suggests that evaluators may still favor equally qualified men due to prevailing cultural beliefs that women are less competent and have lower status in certain professions than men (Correll and Benard, 2006; Correll and Ridgeway, 2003). Additionally, women may be judged more strictly than men due to the existence of "double standards" in the evaluation between men and women (Foschi, 1996; Foschi et al., 1994; Ridgeway, 2019; Ridgeway and Markus, 2022). These theories all predict gender biases against women in applicant evaluations in the form of a female disadvantage.

In contrast, Crandall and Eshleman's (2003) justification-suppression model argues that even when prejudices exist, they may not be expressed in evaluations, leading to no observable gender biases. Suppression may occur because (i) individuals feel social pressure not to adhere to stereotypes if these are considered "old fashioned" and "unacceptable" (ii) individuals fear repercussions (e.g., by being held accountable), for expressing bias even if they still hold traditional stereotypes, or (iii) prejudice threatens their self-concept (e.g., they rather not see themselves as sexist). As Crandall and Eshleman (2003, p. 421) argue, "People suppress prejudice both to maintain a nonprejudiced appearance and to deny prejudice to themselves and maintain a nonprejudiced self-concept."

It follows that in environments where national affirmative action policies are in place to increase women's representation in academia, there is a greater likelihood of suppressing potential prejudice against women. Recent studies have even reported instances of a female advantage in evaluation processes (Card et al., 2020, 2023; Ceci et al., 2023), suggesting that suppression can sometimes extend beyond simply eliminating discrimination. Such instances are particularly pronounced in settings where longstanding gender equality initiatives actively challenge conventional group stereotypes in recruitment practices (Leslie et al., 2017).

By relating observed gender biases to what the above theoretical accounts predict, we can only indirectly infer possible underlying mechanisms. We interpret findings of a male advantage as support for the statistical discrimination and status theory frameworks and findings of no gender bias or a female advantage as support for the justification-suppression model.

The gender norms and policies enacted in German and Italian academia could influence these processes of suppression, even leading to a female advantage, albeit with differing impacts on the two evaluation dimensions of "perceived qualifications" and "invitation propensity" outlined above. Due to national differences in gender policies and normative climates, we formulated country-specific hypotheses. Our theoretical framework involved comparing these country-specific hypotheses to explore the potential dependence of gender prejudice suppression on contextual factors, as posited by the justification-suppression model and consistent with recent studies reporting a female advantage. This comparative analysis is exploratory in nature due to the inability to randomly manipulate gender norms and policies. Furthermore, because the two countries serve as case studies with different gender policy approaches, they may also differ in unaccounted aspects that are beyond the scope of this study.

As discussed above, Germany has well-established gender equality policies in academia and beyond. Yet, they lack a clear definition of "excellence" and "merit," which may, according to statistical discrimination and status characteristics theories, leave room for the expression of prejudice against women's qualifications, potentially resulting in fewer invitations to interviews. However, we argue that these theories overlook the context of relatively progressive gender norms and extensive gender equality policies that have been in place in German academia for several decades, which may increase evaluators' sensitivity to potential gender biases in both qualification assessments and invitation decisions (Marini and Meschitti, 2018). Thus, in line with the justification-suppression model, evaluators in Germany may not express prejudice against women, resulting in no observable bias. Additionally, they may even exhibit bias in women's favor, possibly to signal alignment with the university's values of gender equality, the university's goal of increasing the proportion of female professors, and adherence to gender-preferential selection (Crandall and Eshleman, 2003; Henningsen et al., 2022).

This advantage for women may be more pronounced for invitation propensity than for perceived qualifications because the gender-preferential selection law targets hiring selection and explicitly requires favoring equally qualified women. Evaluators, whether externally or internally motivated, may adhere to this gender-preferential selection policy. This policy may also indirectly impact the ratings of "perceived qualification," albeit to a lesser degree compared to invitations. That is, a female advantage may be observed as respondents seek to self-justify their higher invitation ratings as corresponding to the applicant's "merits" (performance).

Based on these considerations, we hypothesize for Germany.

- A female advantage for invitation propensity, and either no gender bias or a comparatively smaller female advantage for perceived qualification (G-H1).

Conversely, Italy is characterized by a gender-neutral standardization approach in academia, a less supportive normative and institutional gender climate, and greater autonomy for selection committees in the second stage of hiring. On the one hand, the standardization approach may suppress the expression of gender prejudice because the supposedly merit-based standardized criteria for evaluating publication output are expected to reduce ambiguity about who is deemed as being "qualified." Thus, if applied "correctly," the neutral approach could result in no observable bias against women in evaluating applicants' qualifications. However, because the approach is gender-neutral, we do not expect bias reversal in favor of women. On the other hand, the greater autonomy of selection committees in the second stage of hiring—combined with a lack of proactive gender policies and a less supportive normative gender climate characterized by a stronger adherence to traditional gender norms—may enable the expression of gender prejudice to influence decisions on whom to invite for an interview. Thus, even if evaluators adhere to "merit-based" standardized criteria and perceive objectively comparable men and women as equally qualified, they may still express bias against women regarding their employability. In sum, while the gender-neutral standardization approach may suppress expressions of gender prejudice in perceived qualifications, it may not provide adequate motivation for evaluators to suppress their prejudice in invitation decisions. Thus, we hypothesize for Italy.

- A female disadvantage for invitation propensity, but no gender bias for perceived qualifications (I–H1).

Thus, support for these hypotheses would suggest that both country approaches and context are successful in suppressing expressions of prejudice in evaluations of *perceived qualifications*. In contrast, their success in suppressing expressions of prejudice in evaluations of *invitation propensity* may differ.

### 3.2. Gender biases in the presence of ambiguous signals

The gender equality strategies of both Germany and Italy may lead to different predictions when the quality and productivity signals of applicants are *ambiguous*. According to statistical discrimination theory (Arrow, 1973), performance expectations are likely to be less gender-biased when evaluators have more specific information about an applicant's skills and abilities. However, as the cues become more ambiguous, evaluators are more inclined to rely on gender stereotypes and status characteristics such as gender (Correll and Ridgeway, 2003). As discussed above, both the German and Italian strategies may be effective in suppressing potential gender prejudice if applicants' qualification profiles are unambiguous in terms of both quantity and quality. While quantities are easily measurable and can be fairly accounted for, the same does not apply to quality. Recent studies have shown that for collaborative academic outputs in the field of economics, female co-authors receive less credit than male co-authors (for Italy, see Gerxhani et al., 2023; for the US, see Sarsons, 2017; Sarsons et al., 2021). This finding suggests that the more ambiguous an individual's contribution to collaborative work is in terms of content, the more room there is for status discrimination by evaluators. A meta-study conducted by Koch, D'Mello, and Sackett (2015, p. 139) supports this conclusion, indicating that information regarding an applicant's performance

"must be highly diagnostic to counteract stereotypes."

How could ambiguous signals influence evaluations of applicants in Germany and Italy? If an applicant's research output primarily consists of collaborative work, it may be less clear how much and what each applicant has contributed. This ambiguity can affect evaluations of both male and female co-authors (as a co-authorship penalty), but it tends to affect women more severely due to the potential prejudice that female co-authors are less competent and, therefore, must have contributed less than their male counterparts as observed in previous literature (e.g. Sarsons et al., 2021). Consequently, in such ambiguous situations, even Germany's gender-equality policies and normative approaches may not be able to suppress expressions of prejudice, resulting in a larger penalty for women. However, the observed gender difference in invitations may be weaker due to compliance with the "compensating" gender-preferential selection rules. Thus, for Germany, we hypothesize.

- A larger co-authorship penalty for women than men for both perceived qualifications and invitation propensity (G-H2a), but
- A larger gender difference in the co-authorship penalty for perceived qualifications than for invitation propensity (G-H2b).

The more indirect standardization strategy in Italy is designed to ensure a "meritocratic" assessment of applicants' productivity as long as that productivity is easily measurable and fairly accounted for (e.g., through the number of publications). However, this strategy may fail when meritocratic assessments rely on more ambiguous signals of productivity, such as contributions to collaborative works. Additionally, the greater autonomy of selection committees in the second stage of hiring—combined with less progressive gender norms and the absence of proactive gender policies—may create further opportunities for expressing biases against women. This may be particularly true when assessing ambiguous co-authorship signals, affecting both perceived qualifications and invitation propensity. Therefore, for Italy, we hypothesize.

- A larger co-authorship penalty for women than men for both perceived qualifications and invitation propensity (I–H2).

Thus, expectations for both countries are equal regarding gender bias in co-authorship ratings (G-H2a and I–H2); however, differences between the two dimensions are only expected for Germany (G-H2b).

Ambiguity in career *commitment* is also an important factor when evaluating applicants (Correll and Benard, 2006). The post-doctoral stage and transition to assistant professorships often coincide with the transition to parenthood. Recent studies have shown that mothers tend to publish less than fathers (e.g., Morgan et al., 2021), which can significantly lower their chances of being appointed. However, motherhood (in contrast to fatherhood) may also generate a penalty beyond this productivity difference due to gender stereotypes associated with parenthood. According to status characteristics theory, motherhood may be considered a devalued status characteristic in workplace settings such as academia because normative role expectations for motherhood contradict prevailing "ideal worker" norms of highly committed employees (Ridgeway and Correll, 2004). Thus, mothers may be evaluated more harshly than equally qualified nonmothers and men (regardless of whether they have children). Studies have corroborated that parental status indeed operates as a status characteristic: Evaluators often believe that mothers are less committed to work and consider them less competent, while fathers are perceived as being more committed (i.e., productive with less "time on task") compared to nonfathers (e.g., Correll et al., 2007; Cuddy et al., 2004). As a result, family responsibilities may send negative signals about mothers, leading to a motherhood penalty, while signaling higher productivity for fathers, producing a fatherhood premium.

In our study, we used parental leave as an indicator of parenthood status. How evaluators interpret parental leave as an ambiguous signal of an applicant's career commitment may vary depending on the country's normative climate, gender policies, and parental leave policies, which define what is mandatory and considered typical. A shorter-than-mandatory or typical parental leave may be interpreted as a sign of high work commitment, while a longer leave may be perceived as a violation of the ideal worker norm, potentially reinforcing stereotypes of lower work commitment (Weisshaar, 2018). Interestingly, a German study found that mothers who deviated from gender norms by taking only the mandatory two months of parental leave after childbirth (and, thus, "acting like men") were judged more negatively than those who took longer leaves, while fathers' leave duration did not influence evaluations (Hipp, 2018).

In both countries, parental leave regulations for mothers are relatively generous (OECD, 2022, p. 3). In Germany, female employees receive 14 weeks of mandatory maternity leave at 100% of their salaries. Paid parental leave provides a parental allowance of around 67% of the previous net monthly income for up to 12 months, extendable to 14 months if each parent takes at least two months. In Italy, female employees are entitled to 22 weeks of compulsory leave with 80% of their salary, and parental leave (at 30% of one's salary) can be shared between parents or taken by either for up to 26 weeks. Thus, parental leave regulations are more generous in Germany than in Italy.

Moreover, in Germany, guidelines for appointment committees explicitly request that childcare responsibilities be considered when evaluating applicants' performance and productivity, regardless of gender. This policy of "accounting for children" may reduce the motherhood penalty by increasing awareness of potential biases and counterbalancing differences in productivity. Additionally, it may signal higher productivity due to less "time on task" if applicants who have taken parental leave have same level of output, such as the same number of publications, as those without leave. As a result, evaluators may not only suppress expressions of prejudice but could also favor equally productive applicants with parental leave. Thus, contrary to status characteristics theories but in line with the justification-suppression model, we do not anticipate finding gender differences in Germany. Instead, we hypothesize.

– A parental leave premium for both women and men (compared to equally qualified applicants without leave) for both perceived qualifications and invitation propensity (G-H3).

Since 2016, the Italian habilitation regulation has required adjustments to the number of publications needed for academic qualification, taking into account documented mandatory maternity leave for women and actual leave taken by both men and women. Consequently, the required publication thresholds are lowered for parents. This regulation reduces the likelihood of arbitrary considerations of childcare responsibilities negatively impacting the evaluation of qualifications. Additionally, these adjustments may even favor applicants with parental leave who have the same level of productivity (e.g., the same number of publications) as those without leave. However, appointment committees are not required to consider childcare responsibilities in the second stage of hiring, that is, interview invitations. The absence of official policies to "discount" childcare responsibilities during the later stages of the selection process, coupled with traditional views of motherhood reflected in prevailing gender norms (see Section 2) and a lack of direct gender equality initiatives during the appointment process, may fail to suppress expressions of prejudice against mothers in invitations. Thus, in contrast to Germany, we expect to find in Italy.

– Either no penalty or a parental leave premium for both women and men (compared to equally qualified applicants without leave) for perceived qualifications, but a penalty for mothers regarding invitation propensity (I–H3).

## 4. Data and methods

### 4.1. Sample

To test our hypotheses, we conducted factorial survey experiments (also known as vignette studies) in Germany and Italy. The respondents were gatekeepers, that is, professors who typically serve as members of appointment committees in the two countries. In accordance with the country-specific eligibility criteria for committee membership, the German sample included assistant, associate, and full professors in the fields of economics, political science, and social sciences, while the Italian sample included only associate and full professors of the same disciplines. Therefore, in our analyses, we controlled for respondents' professorship rank. We deliberately chose this sampling strategy to increase the external validity of our study, aiming to simulate the hiring process in both countries, where prospective faculty are evaluated by current faculty members (see Ceci, 2018).[6] The survey experiments in the two countries were harmonized by virtue of being jointly designed but were administered one year apart for logistical reasons related to data collection. The German survey experiment was administered in June/July 2020 using LimeSurvey, while the Italian survey experiment was carried out in June 2019 using Qualtrics.

In Germany, the gross sample was collected from all public universities' web pages in the selected disciplines and departments.[7] The gross sample included 4117 addresses of professors, out of which 990 professors participated in the survey (response rate of 24%).[8] Because the websites rarely included information on professorship rank, this information was collected at the end of the survey, which resulted in missing values due to item non-response. In our analyses, we excluded 112 respondents with missing information on their professorship rank, 72 respondents who were emeritus or retired, 74 who held another type of professorship (e.g., temporary substitute professor or visiting professor), and 43 with missing values for the dependent variables. This resulted in an analytical sample of 689 respondents from Germany.

In Italy, the survey was sent to 1474 associate and full professors who were randomly sampled from the official registry of professors of the Italian Ministry of education (MIUR) in April 2018. Of these, 266 professors participated in the survey (response rate of 18%). Fifty-five respondents were excluded due to missing values for the dependent variables. Thus, the total analytical sample for Italy consisted of 211 respondents.

These exclusions do not compromise the experimental design, as shown by the correlation matrix of the vignette dimensions (i.e., all dimensions remained uncorrelated and thus independent) and the equal distributions of the vignette levels in both the German and Italian analytical samples (see Appendix, Tables A2 and A3).

Table 1 provides additional details on the samples and response rates. In both countries, female professors and those from political and social sciences were more likely to participate in the survey. The higher share of female participants was not influenced by the perception that our study was gender-related, as the invitation and introduction to both surveys referred to "recruitment criteria for professorships" without mentioning gender. Moreover, our between-subject design regarding applicants' gender made it difficult to detect our gender-related research question, as explained below. Non-response may have been influenced by factors beyond our control, such as time constraints (which could typically favor men) or a general aversion to surveys. To facilitate the country comparison, we controlled for respondents' gender and disciplinary field.

---

[6] Doing so meant that the German sample was, on average, younger than the Italian sample. It could be argued that younger individuals may hold fewer traditional gender stereotypes than older individuals, whether due to cohort effects or age-related factors. Therefore, we also conducted regressions that included academic age (i.e., the number of years since PhD graduation). The coefficients for respondents' age were found to be very small and not statistically significant in both countries. Moreover, the results for applicants' gender remained the same. For Germany, we also estimated the regressions excluding assistant professors, and the results are congruent to the full sample (see Online Supplement, Table S3).

[7] Universities of applied sciences were not included because they typically do not have assistant professors.

[8] Due to the experimental design, potential self-selection into the sample did not affect the internal validity of the study but it may reduce its external validity.

**Table 1**
Samples and response rates for Germany and Italy.

|  | Gross | Response rate | Realized sample | Analytical subsample |
|---|---|---|---|---|
| **Germany** |  |  |  |  |
| Total number of respondents | 4117 | 24.0 | 990 | 689 |
| Male professors | 3179 | 22.3 | 710 | 473 |
| Female professors | 938 | 29.8 | 280 | 216 |
| % of female professors | 22.8 |  | 28.3 | 31.3 |
| Number of vignettes |  |  |  | 4106 |
| Share of responding professors from |  |  |  |  |
|    Economics (%) | 62.6 | 18.0 | 46.9 | 48.8 |
|    Political & social sciences (%) | 37.4 | 34.2 | 53.1 | 51.2 |
| **Italy** |  |  |  |  |
| Total number of respondents | 1474 | 18.05 | 266 | 211 |
| Male professors | 945 | 15.8 | 149 | 120 |
| Female professors | 529 | 22.1 | 117 | 91 |
| % of female professors | 35.9 |  | 44.0 | 43.1 |
| Number of vignettes |  |  |  | 1266 |
| Share of responding professors from |  |  |  |  |
|    Economics (%) | 58.4 | 13.1 | 42.5 | 42.7 |
|    Political & social sciences (%) | 41.6 | 25.0 | 57.5 | 57.4 |

## 4.2. Experimental design

We asked participants to evaluate descriptions of fictional applicants for assistant professorships in Germany and Italy on two dimensions: (1) the likelihood of considering the applicant as qualified for an assistant professorship (perceived qualification) and (2) the likelihood of inviting the applicant for an interview (invitation propensity), using a scale from 1 (lowest) to 7 (highest). Participants evaluated the applicants' profiles relative to their respective disciplines. The vignette dimensions (or dimensions for applicants' profiles) of interest for this study included the applicant's gender, the authorship characteristics of the applicant's scientific publications (i.e., all co-authored vs. mostly co-authored vs. mostly single-authored), and whether or not the applicant reported parental (maternity/paternity) leave (see Table 2). All dimensions included in the factorial survey are reported in the Appendix, Table A3. Examples of the vignettes and the opening letters are provided in the Appendix, Tables A4a and b.

We made some adjustments when implementing the experiments to accommodate national differences in hiring procedures. A major difference concerned the number of publications attributed to the applicants in the experiment, which was treated as a fixed characteristic in the vignettes. In Italy, the number of publications was predefined and uniform across the three fields, reflecting the current requirements of the national habilitation process (as discussed above).[9] In contrast, Germany lacks specific standards regarding the required number of publications. Therefore, we asked respondents at the beginning of the survey about their opinions on the approximate number of peer-reviewed publications an applicant should have in their discipline to be considered for an assistant professorship four years after completing their PhD. This respondent-specific number of publications was included as a fixed (stable) characteristic in the applicants' profiles in all vignettes evaluated by the respective respondents and was controlled for in the analyses.[10]

In both countries, applicants' profiles varied regarding the proportions of single- and co-authored peer-reviewed publications (see Table 2). For Italy, we operationalized this variation using a predefined number of publications, while for Germany, we used the proportions of single-authored and co-authored publications, harmonizing the shares according to the distribution implemented in the Italian vignettes. Additionally, based on country-specific parental leave regulations, we set the length of parental leave at five months in Italy (i.e., the mandatory period for women and the most typical duration in Italian academia, given the low payment for longer leaves) and six months in Germany (where longer leaves than the mandatory period are more common).

Finally, to provide more realistic descriptions, we also added the following information to the vignettes as fixed characteristics. For Germany, all applicants were described as having completed their PhD with the grade of magna cum laude (very good) because German PhD theses are graded. Applicants also had four years of experience since completing their PhDs, which is typical for assistant professorship applications and qualifies them for assistant professorship appointments across all federal states. In the Italian vignettes, we also indicated that applicants had four years of experience after their PhDs for the untenured assistant professor positions and six years for the tenured positions. Additionally, we noted the presence of prior funding for their own research to realistically depict the formal and informal requirements for a successful application for an assistant professorship.

---

[9] According to the national habilitation rules, the vignettes in the Italian survey experiment differentiated between the number of publications: five for non-tenure-track positions and ten for tenure-track positions (type of position is included in all regressions).

[10] Respondents who did not answer this question or who reported less than three publications were shown applicants' profiles with three publications in order to generate meaningful vignettes, also with respect to the publication record levels (co-authorships).

**Table 2**
Vignette dimensions of applicants' profile of interest for the study.

| Dimensions | Levels (Germany) | Levels (Italy) |
|---|---|---|
| Applicant's gender (between-subject design, see below) | Female [first + last name]<br>Male [first + last name] | Female [first + last name]<br>Male [first + last name] |
| Publication record | - Mostly single-authored publications (70% single and 30% co-authored)<br>- Mostly co-authored publications (30% single and 70% co-authored)<br>- All co-authored publications | *When applying for a tenure-track position:*<br><br>- Mostly single-authored publications (7 single and 3 co-authored)<br>- Mostly co-authored publication (3 single and 7 co-authored)<br>- All co-authored publications<br>*When applying for a non-tenure-track position:*<br>- Mostly single-authored publications (4 single and 1 co-authored)<br>- Mostly co-authored publications (1 single and 4 co-authored)<br>- All co-authored publications |
| Parental leave | –<br>6 months of parental leave | –<br>5 months of parental leave |

Note: The applicants' profiles included additional dimensions that were not of interest to this study but were included as control variables. These dimensions were research collaborations (with renowned scholars or scholars at the same career level), tenure or non-tenure-track positions, and the acquisition of a research grant (Germany). See Appendix, Table A3.

The combination of all vignette dimensions yielded a universe of 24 different combinations in Italy and 48 combinations in Germany.[11] The vignettes were grouped into four blocks of six vignettes in Italy and eight blocks of six vignettes in Germany. These blocks were then randomly assigned to respondents (within economics and political/social sciences, respectively). We relied on the d-optimal blocking approach to generate the blocks and randomized the order of the vignettes within each block. This procedure ensured that none of the dimensions were confounded with the blocks and enabled us to estimate main and two-way interaction effects between vignette dimensions without the risk of confounding (Su and Steiner, 2020; Treischl and Wolbring, 2022). To avoid censoring the rating, respondents were able to go back and forth between the six vignettes (Auspurg and Hinz, 2015, p. 214). Due to the combination of sample size and six vignettes per respondent, the total number of evaluated vignettes in the analysis was 4106 in Germany and 1266 in Italy.[12]

Like Carlsson et al. (2021) and Ceci (2018), we employed a between-subject design for applicants' gender (where respondents evaluated only male vignettes or only female vignettes) alongside a within-subject design for all other vignette dimensions. The between-subject design is demonstrably effective in avoiding the possibility of respondents detecting the study's purpose (here: gender inequality), and thus effectively reduces the social desirability bias (Walzenbach, 2019). As shown by Koch, D'Mello, and Sackett (2015, p. 131, p. 138) in their meta-analysis, results generated from rating only male or female profiles do not differ from those obtained through competitive ratings of both genders.

*4.3. Analytical strategy*

Corresponding to our nested data structure, we estimated multi-level linear models with random intercepts to investigate the effects of the vignette dimensions on the likelihood of an applicant being considered qualified and invited for an interview (rated on a scale from 1 to 7) (Auspurg and Hinz, 2015, p. 92).[13] The unit of analysis was the evaluated vignettes. We conducted the estimations separately for Germany and Italy, using discipline-pooled analyses. The selected disciplines were relatively homogeneous in their requirements, as evidenced by the Italian national habilitation (i.e., applicants require a similar number of high-quality publications to qualify for the habilitation). In Germany, respondents in both fields indicated that, on average, six peer-reviewed publications were necessary for an applicant to have a chance of being appointed.

To test the first hypotheses (G-H1 and I–H1), we estimated the effect of applicants' gender on evaluations of perceived qualifications and invitation probability. For the remaining hypotheses, we included interaction terms: applicant's gender * type of publication record, and applicant's gender * parental leave. All models controlled for respondents' characteristics, including their gender,

---

[11] Including one more dimension in Germany does not change the magnitude of the gender effect for two main reasons: (1) the experimental dimensions were designed to be independent of one another (see also the correlation matrix in Appendix, Table A2), and (2) the interaction effect between gender and research grants was not significant and almost zero.

[12] A priori sample size was calculated using a full factorial design that accounted for both the main (vignette) effects and two-way interactions between the main vignette effects. This calculation set the type I error alpha at 0.05 and the type II error at 0.2, as per convention. The minimum sample size needed to detect a large effect related to gender, given the universe of 24 vignettes and the combination of between-subject and within-subject designs, was estimated to be approximately 50 participants. Sensitivity power analyses indicated that the sample size needed to detect medium effects was 180–200 participants.

[13] Selection into "treatment" by unobserved characteristics of respondents is not possible due to our randomized vignette assignment (see above).

**Table 3**

Descriptive results for the dependent variables, Germany and Italy.

| | Germany | | | Italy | | |
|---|---|---|---|---|---|---|
| | Observations[a] | Mean | SD | Observations[a] | Mean | SD |
| **All respondents** | | | | | | |
|   Invitation propensity | 4102 | 5.36 | 1.43 | 1266 | 5.85 | 1.29 |
|   Perceived qualification | 4106 | 5.29 | 1.42 | 1266 | 5.34 | 1.32 |
| **Female applicants** | | | | | | |
|   Invitation propensity | 2050 | 5.48 | 1.42 | 598 | 5.86 | 1.29 |
|   Perceived qualification | 2059 | 5.41 | 1.42 | 598 | 5.33 | 1.31 |
| **Male applicants** | | | | | | |
|   Invitation propensity | 2052 | 5.23 | 1.42 | 628 | 5.84 | 1.29 |
|   Perceived qualification | 2047 | 5.18 | 1.42 | 628 | 5.34 | 1.34 |
| **Female respondents** | | | | | | |
|   Invitation propensity | 1282 | 5.37 | 1.48 | 525 | 5.80 | 1.35 |
|   Perceived qualification | 1286 | 5.33 | 1.41 | 525 | 5.24 | 1.38 |
| **Male respondents** | | | | | | |
|   Invitation propensity | 2820 | 5.35 | 1.40 | 701 | 5.89 | 1.24 |
|   Perceived qualification | 2820 | 5.28 | 1.43 | 701 | 5.42 | 1.28 |

Original scale: 1 (lowest) to 7 (highest).

[a] Number of vignettes; SD = standard deviation.

disciplinary area (economics or political/social sciences), and professorship rank, as well as design variables such as the order of vignettes and equal evaluations of vignettes by respondents (see Online Supplement, Table S1a/b).

To facilitate comparisons of effect sizes across the samples, we standardized the outcome variables, as the evaluators in the two countries may differ in their rating behavior independent of gender bias; therefore, we present the results as a proportion of one standard deviation. We also estimated regressions using the original scale (see Online Supplement, Table S2a/b). To test the statistical significance of differences in effect sizes for the two evaluation dimensions within Germany (see Hypotheses G-H1 and G-H2b) and, exploratively, for differences between countries, we used z-tests assessed with the usual t-statistics, as outlined by Clogg et al. (1995; see also Paternoster et al., 1998). Given that our hypotheses were directed, we used a one-tailed test.

## 5. Results

Table 3 presents the descriptive findings for the two countries. The mean rating of invitation propensity was higher in Italy (5.85) than in Germany (5.36), while the mean rating of perceived qualifications was roughly similar across both countries (5.34 in Italy vs. 5.29 in Germany).[14] In Germany, women received slightly higher ratings on average than men for both invitation propensity and perceived qualifications, while there were no differences in Italy. Within each country, the mean ratings given by female and male respondents were comparable for both outcome variables. These country differences may be due to differences in the sample compositions (see Table 1) or differences in general rating behavior. Therefore, we estimated regressions that included respondents' characteristics and standardized outcome variables.

Table 4 presents the regression results related to Hypotheses G-H1 and I–H1. Our analyses show that German professors rated women higher than men on both invitation propensity and perceived qualification by 18.6% and 17.3% of one SD, respectively, equating to 0.266 and 0.246 rating points, respectively (see Online Supplement, Table S2a). These results support the first part of Hypothesis G-H1, which predicted a female advantage for invitation propensity. However, they do not support the second part, which predicted no bias or a smaller female advantage for perceived qualifications compared to invitation propensity in Germany. Women seem to have had a similar advantage in both evaluation dimensions in Germany (z-tests are also not significant).

In contrast, there was neither a female advantage nor disadvantage in Italy for either outcome. These findings contradict the first part of Hypothesis I–H1, which predicted a female disadvantage in invitation propensity in Italy but support the second part (no gender bias in perceived qualifications). Finally, country differences in gender bias were only tentatively supported for both invitation probability and perceived qualifications, with respective z-tests significant at the 10% level.

Additional vignette dimensions included were research collaborations, tenure- or non-tenure-track positions, and research grants (Germany). Additional controls included respondents' gender, rank, discipline, order of vignettes, a dummy for equal evaluations of all vignettes, and respondent-specific number of required publications (Germany).

Regarding the ambiguous signals of applicant productivity due to co-authorships, Table 4 reveals that in both countries, applicants with all co-authored papers received lower ratings, while those with mostly single-authored papers received higher scores. In Germany, the negative ratings applied to both outcomes. In Italy, however, co-authorship had a greater negative impact on perceived qualifications than on invitation propensity, which contrasts with the expected standardization approach for accrediting competence and eligibility. The difference between invitation propensity and perceived qualifications in Italy was statistically significant.

Concerning parental leave, we found modest but notable positive effects in both countries. Since applicants' profiles were designed

---

[14] The distributions are skewed toward higher ratings because all applicants are formally eligible for assistant professor positions.

**Table 4**

**Multi-level linear regressions with random intercept: Invitation propensity and perceived qualifications for Germany and Italy** (coefficients as a proportion of SD).

| | Germany | | Italy | |
|---|---|---|---|---|
| | Invitation propensity | Perceived qualification | Invitation propensity | Perceived qualification |
| Applicant's gender (ref: male applicant): female applicant | 0.186** | 0.173** | −0.015 | −0.015 |
| | (0.064) | (0.063) | (0.106) | (0.104) |
| Publication record (ref: mostly co-authored) | | | | |
| All co-authored | −0.304** (0.019) | −0.319** (0.019) | −0.229** (0.049) | −0.326** (0.056) |
| Mostly single-authored | 0.228** (0.019) | 0.224** (0.019) | 0.177** (0.046) | 0.258** (0.053) |
| With parental leave (ref: w/o leave) | 0.098** (0.016) | 0.102** (0.016) | 0.029 (0.036) | 0.094* (0.042) |

Standard errors in parentheses, **$p < 0.01$, *$p < 0.05$ (two-sided tests); SD = standard deviation. Only coefficients of interest are included; full models are presented in Online Supplement, Table S1a/b.
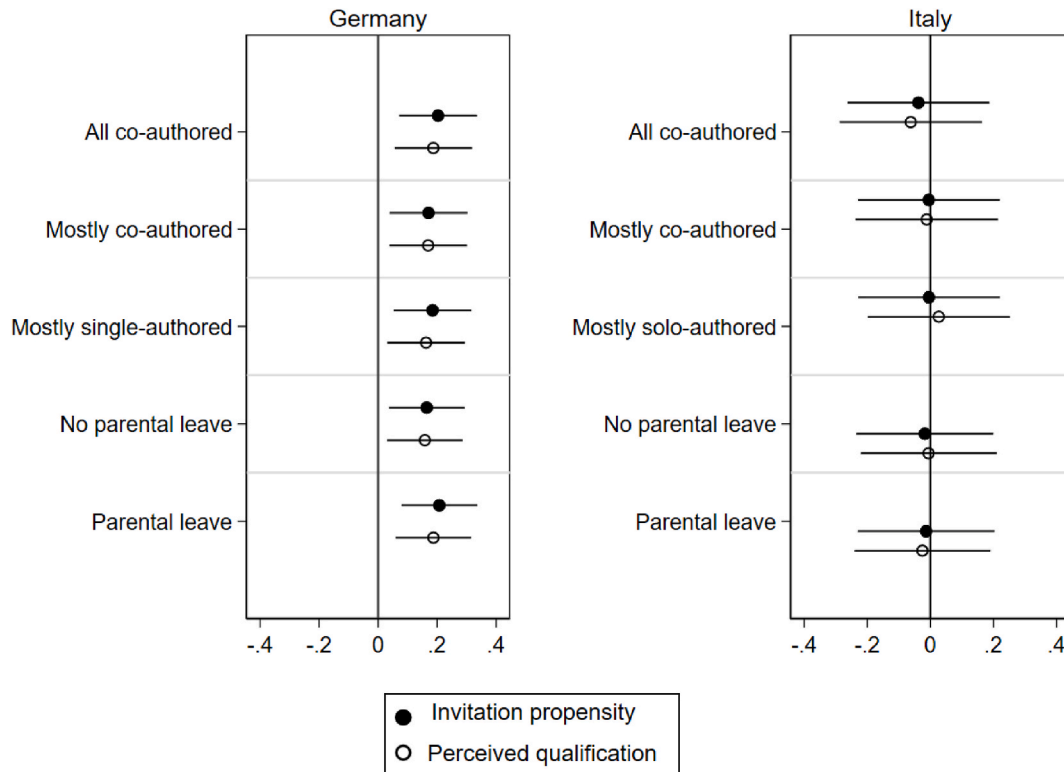


**Fig. 1.** AMEs of applicants' gender (women compared to men) by type of publication record and parental leave (as a proportion of SD, 95% CI).

Note: Positive values indicate a female advantage, while negative values indicate a male advantage. Based on regressions reported in Online Supplement, Table S1a/b.

to be equal in the standard measures of "productivity" (such as the number of publications), we expected that applicants with parental leave who were as productive as those without would get a "premium." For Germany, we observed a statistically significant advantage for applicants with parental leave for both perceived qualifications and invitation propensity. In Italy, this premium was only noted for perceived qualifications. Accordingly, the country difference in the magnitude of the parental leave effect was only statistically significant for invitation propensity. These findings are in line with the German policy that requires family time to be considered when evaluating applications, while in Italy, parental leave is modestly considered in the first stage of habilitation but not in the second stage of selection committees' decisions.

Our primary interest in this vein concerned gender biases related to ambiguous signals of productivity, as measured by co-authorship (Hypotheses G-H2a/b and I–H2), and career commitment, as indicated by parental leave (Hypotheses G-H3 and I–H3). The results for these measures are presented in Fig. 1 (for the full models, see Online Supplement, Table S1a/b). In both countries, none of the interactions between applicants' gender and the type of publication record or parental leave were statistically significant (i.e., the AMEs were equal across the vignette dimensions). Thus, none of the respective hypotheses regarding a female co-authorship penalty were supported. For Italy, no female disadvantage was observed when there was greater ambiguity. In Germany, female

**Table 5**

**Multi-level linear regressions with random intercept: invitation propensity and perceived qualifications, separate models for respondent's gender, Germany and Italy** (coefficients as a proportion of SD).

| | Invitation propensity | | | | | | Perceived qualification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Male p. | Female p. | Male p. | Female p. | Male p. | Female p. | Male p. | Female p. | Male p. | Female p. | Male p. | Female p. |
| **Germany** | | | | | | | | | | | | |
| Applicant's gender (ref: male): female | 0.167* | 0.249* | 0.154 | 0.220 | 0.143 | 0.231* | 0.137 | 0.255* | 0.145 | 0.220 | 0.115 | 0.254* |
| | (0.076) | (0.113) | (0.081) | (0.120) | (0.078) | (0.116) | (0.077) | (0.109) | (0.081) | (0.117) | (0.079) | (0.113) |
| *Publication record (ref: mostly co-authored)* | | | | | | | | | | | | |
| All co-authored | −0.335** | −0.241** | −0.345** | −0.274** | −0.335** | −0.241** | −0.348** | −0.254** | −0.341** | −0.302** | −0.349** | −0.254** |
| | (0.023) | (0.034) | (0.033) | (0.047) | (0.023) | (0.034) | (0.022) | (0.036) | (0.032) | (0.049) | (0.022) | (0.036) |
| Mostly single-authored | 0.214** | 0.256** | 0.205** | 0.250** | 0.214** | 0.256** | 0.208** | 0.259** | 0.212** | 0.258** | 0.208** | 0.259** |
| | (0.023) | (0.034) | (0.033) | (0.047) | (0.023) | (0.034) | (0.022) | (0.036) | (0.032) | (0.049) | (0.022) | (0.036) |
| With parental leave (ref: w/t leave) | 0.079** | 0.135** | 0.079** | 0.135** | 0.054* | 0.118** | 0.075** | 0.160** | 0.075** | 0.160** | 0.052 | 0.159** |
| | (0.019) | (0.028) | (0.019) | (0.028) | (0.027) | (0.039) | (0.019) | (0.030) | (0.019) | (0.030) | (0.027) | (0.040) |
| Female*all co-authored | | | 0.020 | 0.072 | | | | | −0.014 | 0.103 | | |
| | | | (0.045) | (0.069) | | | | | (0.045) | (0.072) | | |
| Female*mostly single-authored | | | 0.017 | 0.014 | | | | | −0.008 | 0.001 | | |
| | | | (0.045) | (0.069) | | | | | (0.045) | (0.072) | | |
| Female*with parental leave | | | | | 0.048 | 0.036 | | | | | 0.044 | 0.003 |
| | | | | | (0.037) | (0.057) | | | | | (0.037) | (0.059) |
| **Italy** | | | | | | | | | | | | |
| Applicant's gender (ref: male): female | −0.074 | 0.068 | 0.001 | −0.012 | −0.059 | 0.050 | −0.044 | 0.020 | 0.090 | −0.150 | −0.034 | 0.031 |
| | (0.132) | (0.174) | (0.144) | (0.184) | (0.138) | (0.179) | (0.130) | (0.172) | (0.144) | (0.189) | (0.137) | (0.181) |
| *Publication record (ref: mostly co-authored)* | | | | | | | | | | | | |
| All co-authored | −0.247** | −0.209* | −0.189* | −0.254** | −0.248** | −0.208** | −0.317** | −0.347** | −0.201* | −0.450** | −0.317** | −0.347** |
| | (0.069) | (0.067) | (0.084) | (0.087) | (0.069) | (0.067) | (0.074) | (0.084) | (0.090) | (0.110) | (0.074) | (0.084) |
| Mostly single-authored | 0.235** | 0.104 | 0.284** | 0.028 | 0.235** | 0.104 | 0.354** | 0.128 | 0.431** | −0.028 | 0.354** | 0.128 |
| | (0.064) | (0.065) | (0.080) | (0.085) | (0.064) | (0.065) | (0.069) | (0.082) | (0.086) | (0.106) | (0.069) | (0.082) |
| With parental leave (ref: w/t leave) | 0.005 | 0.051 | 0.006 | 0.049 | 0.019 | 0.033 | 0.082 | 0.094 | 0.083 | 0.090 | 0.091 | 0.106 |
| | (0.050) | (0.052) | (0.050) | (0.052) | (0.063) | (0.069) | (0.054) | (0.066) | (0.054) | (0.065) | (0.068) | (0.087) |
| Female*all co-authored | | | −0.122 | 0.090 | | | | | −0.243* | 0.205 | | |
| | | | (0.101) | (0.106) | | | | | (0.108) | (0.132) | | |
| Female*mostly single-authored | | | −0.102 | 0.147 | | | | | −0.158 | 0.300* | | |
| | | | (0.101) | (0.106) | | | | | (0.109) | (0.132) | | |
| Female*with parental leave | | | | | −0.029 | 0.035 | | | | | −0.020 | −0.024 |
| | | | | | (0.083) | (0.086) | | | | | (0.089) | (0.108) |

Male p. = male professors, female p. = female professors. Standard errors in parentheses, **p < 0.01, *p < 0.05 (two-sided tests). Only coefficients of interest are included (full models available upon request from the authors). Two-way interactions "respondent gender * vignette dimensions" in pooled models. Germany: statistically significant differences for respondents' gender*all co-authored publications and respondents' gender*parental leave. Italy: no two-way interactions were statistically significant.

applicants received higher scores than their male counterparts in both evaluations, regardless of the type of publication record. Concerning parental leave, the results support Hypothesis G-H3, which predicted a parental leave premium for both mothers and fathers in Germany. In Italy, the findings align with the first part of Hypothesis I–H3, which anticipated either no penalty or a parental leave premium related to perceived qualifications for both mothers and fathers. However, the results contradict the second part of the hypothesis, which expected a parental leave penalty in invitation propensity for mothers.[15]

As a sensitivity check, we examined whether the evaluations varied according to respondents' gender (see Table 5). In Germany, women received higher ratings than men from both male and female professors. Moreover, for professors of both genders, none of the interactions between applicants' gender and our two vignette dimensions signaling ambiguity were statistically significant. Thus, these separate analyses for Germany confirm the findings for the whole sample.

In Italy, applicants' gender was also not relevant for the evaluations by either female or male professors (see Table 5). However, the interactions between applicants' gender and co-authored publications differed for male and female professors, with male professors penalizing women more when all their publications were co-authored and female professors penalizing men more when all their publications were co-authored (the difference was significant only at the 10% level). This finding suggests that output standardization may not effectively suppress expressions of gender bias by male evaluators when assessing ambiguous signals of productivity, thereby partially supporting Hypothesis I–H2 regarding the female co-authorship penalty.

## 6. Discussion and conclusions

Discourses on gender inequality in academia have been primarily dominated by explanations for discrimination. Previous research in this respect has largely focused on hiring at the associate and full professor levels. Our study shifts attention to the relatively less explored issue of gender bias in hiring assessments for assistant professorships. Assistant professorships serve as an important stage between lower-level and tenured positions, and biases at this juncture can significantly contribute to the leaky pipeline and the underrepresentation of women in senior academic positions. However, research on assistant professorship hiring remains limited, with existing studies offering inconclusive results. Moreover, no previous study has compared this issue across different national contexts.

We investigated gender bias in two key phases of assistant professor recruitment— perceived qualification and invitation propensity—using harmonized factorial survey experiments. Our sample comprised professors from the fields of economics, political science, and social sciences in Germany and Italy. While statistical discrimination and status characteristic theories predict bias against women in academia, and therefore a female disadvantage, Crandall and Eshleman's (2003) justification-suppression model posits that progressive policy and normative contextual factors may suppress the expression of gender stereotypes or prejudices. By examining two countries with distinct gender norms and policies, we explored how context affects the suppression of prejudice expression in assistant professorship recruitment. Germany, with more progressive norms and direct gender equality measures, contrasts with Italy's more traditional gender norms and gender-neutral policies in academia, offering a unique lens for understanding how different contexts address gender prejudice.

Our findings reveal that, in Germany, female applicants held an advantage in both perceived qualifications and invitation propensity. In contrast, evaluations in Italy showed no advantage or disadvantage for women. This largely aligns with our expectations about the between-country differences resulting from distinct gender norms and equality policies, supporting the justification-suppression model's predictions for observing either no gender bias or a female advantage. An alternative and complementary explanation for this female advantage is that women may be perceived as bringing "additional diversity value" meaning that the working environment benefits from their presence (Schearer et al., 2023, p. 3; see also Chang et al., 2019; Leslie et al., 2017). These two explanations are not mutually exclusive: if gender prejudices exist, diversity values can be one reason for suppressing their expression, and when strongly emphasized, these values may even lead to favoring female applicants in hiring decisions, particularly in the context of women's underrepresentation.

Contrary to our expectations, we did not observe gender bias related to collaborative work. Regardless of the applicant's gender, we found a co-authorship penalty and a single-authorship premium in both countries. As hypothesized, we observed a parental leave premium for both mothers and fathers in both evaluations in Germany and in perceived qualifications in Italy. However, despite the expectation of motherhood penalty for invitation propensity in Italy, we found no evidence of it.

Our results align with more recent studies that have not found gender bias and have even observed a female advantage over the past two decades (see the systematic review by Ceci et al., 2023). The different time periods studied might serve as one of the explanations why our findings differ from the female disadvantage observed in the study by Steinpreis et al. (1999) and by Ooms et al. (2019). During the time of these studies, prejudice and discrimination against women may have been more prevalent and openly expressed. However, decades of efforts and policies aimed at addressing such biases may have contributed to the improvements observed in more recent studies, including ours. Additionally, our study covered all German universities in economics, political science, and social sciences, while Ooms et al. (2019) focused on a single German technical university and examined more male dominated disciplines like mathematics, computer science, natural sciences, and engineering.

In sum, the results from our cross-country study challenge the notion that bias against women in professorship appointments

---

[15] We also estimated interactions between applicant's gender and the other vignette dimensions (collaborations, type of position, and, for Germany, research grants). In Germany, all interactions were insignificant. In Italy, there was a significant interaction effect between gender and the type of professorship for perceived qualifications, indicating that the rating difference between tenure-track and non-tenure-track positions was smaller for male applicants (8.6% of one SD) than for female applicants (15.3%). Results are available upon request.

remains the primary reason for women's underrepresentation in academic hiring despite the existence of affirmative policies. Instead, our findings suggest that in such contexts, even if prejudices exist, their expression is suppressed, and bias may even be reversed in response to gender equality regulations in academia. Our exploratory country comparison also indicates that Germany's more progressive gender norms and policies may be more effective in addressing gender bias and ultimately increasing the share of female assistant professors.

However, gender stereotypes and prejudices may still manifest in later stages of the appointment process, which were not covered in our study. Moreover, addressing the issue as to why women are less likely to apply for professorships in the first place is crucial for increasing the share of female professors. This challenge is complex and multifaceted, with gender biases still observed in areas such as teaching evaluations (Ceci et al., 2023) and authorship crediting (Ross et al., 2022, p. 135). Moreover, the prevailing perception of women's discrimination in faculty appointments may discourage women from applying, thus further contributing to their underrepresentation. This would mean that part of women's underrepresentation results from the Thomas theorem, "If individuals define situations as real, they are real in their consequences," rather than actual discrimination (see Merton, 1995). We hope that our findings, in line with the recent systematic review by Ceci et al. (2023), will raise awareness of the current academic landscape and the potential role of normative and institutional contexts therein, ultimately encouraging more women to apply for assistant professorships.

Some limitations of our study should be acknowledged. First, factorial experiments represent a "low-stakes" scenario in which respondents evaluate hypothetical applicants rather than making actual hiring decisions. This raises concerns about the external validity of our findings, particularly regarding the female advantage observed in Germany. However, official statistics from 2019 support our results, indicating that the share of female applicants for assistant professorships in economics, law, and the social sciences that year was 31%. Among these applicants, 44% of those shortlisted were women, and ultimately, 49% of those appointed were female (GWK, 2020, p. 38, 87). While these statistics could also reflect a combination of bias reversal in favor of women and differences in applicants' qualifications (e.g., men applying with a lower fit for a given position; see Urry, 2015), they nonetheless signal a similarity between real-life hiring and our experimental findings.

Second, our exploratory results are based solely on a policy comparison between two countries. Our findings suggest that a significant proportion of German academics have internalized the goal of increasing women's representation in professorships. Thus, our results hint at Acker's suggestion (Acker, 2006) that when universities' affirmative action policies influence organizational practices, they could be a successful means of counteracting gender biases in assessments and evaluations. However, additional country cases are needed to better understand the role of normative and institutional factors—specifically, under which conditions the expression of potential prejudice is suppressed or not.

Third, our Italian survey was conducted before the COVID-19 pandemic, while the German one was completed shortly after the first lockdown, during a period of online teaching and widespread home office use in German universities. This may have impacted the German study: Fewer professors with children may have participated due to school closures and the resulting increase in caregiving responsibilities. If so, we may be underestimating the parental leave premium and perhaps the female advantage in Germany. On the other hand, lively discussions of gender differences in productivity during this period may have raised awareness of gender issues in German academia, potentially leading us to overestimate the female and parental leave advantage. However, such debates did not occur in isolation but could have resulted from pre-existing gender awareness in German universities.

Finally, our analyses are limited to the fields of economics, political science, and the social sciences. Further research is warranted to examine whether similar processes occur in other academic disciplines and to what extent these findings can be generalized across disciplines.

## CRediT authorship contribution statement

**Klarita Gërxhani:** Writing – review & editing, Writing – original draft, Conceptualization. **Nevena Kulic:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Alessandra Rusconi:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Heike Solga:** Writing – review & editing, Writing – original draft, Conceptualization.

## Research Ethics

The results presented in this article do not allow for deductive disclosure of the survey participants' identities. Participation in the survey was voluntary and participants' confidentiality was protected. The study received ethical approval from the WZB research ethics committee (Nr. 2019/4/73).

## Data availability statement

The data for Italy will be shared upon request to Nevena Kulic. The German data are still being used for a dissertation. They will be shared for replication of this study upon request to Alessandra Rusconi. After completion of the dissertation, the German data of this study will be available from https://doi.org/10.7802/2742.

## Conflicts of interest

The authors declare that they have no conflicts of interest.

## Acknowledgements

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssresearch.2024.103113.

## Appendix

**Table A1**
Summary of key information on studies focusing on different types of professorship appointments (alphabetical order)

| Authors | Country | Academic rank/ level | Disciplines | Data, method, sample size | Observation years |
|---|---|---|---|---|---|
| **Studies on assistant professor appointments** | | | | | |
| Carey et al. (2020) | US | Assistant professor | All | Experimental data, two universities, discipline-pooled analyses, n = 1071 professors, n = 2007 students | 2017 |
| Ceci and Williams (2015); Williams and Ceci (2015) | US | Tenure-track assistant professor | Biology, Economics, Engineering, Psychology | Experimental data, n = 873 prof. From 371 universities/colleges | 2014 |
| Gerxhani et al. (2023) | Italy | Assistant professor (with/without tenure track) | Economics, Humanities, Social Sciences | Experimental data, mainly pooled analyses, n = 395 professors | 2019 |
| Ooms et al. (2019) | Germany, Netherlands | Assistant (& associate, full) professor | Disciplines present in the two universities considered | Observational survey data, one university per country, country- and discipline-pooled analyses, n = 248 academics (holding a PhD) | 2012–2013 |
| Steinpreis et al. (1999) | US | Assistant (& associate) professor | Psychology | Experimental data, n = 238 professors | 1997 |
| **Studies on promotion/hiring of associate/full professors** | | | | | |
| Aksnes et al. (2022) | Norway | Full professor | All | Register data, mainly discipline pooled analyses, n = 19,672 PhD recipients | PhD recipients btw. 1975–2018 |
| Auspurg et al. (2017) | Germany | Associate/full professor | All | Observational data of 235 appointment processes for one university, discipline-pooled analyses | 2001–2013 |
| Bosquet et al. (2019) | France | Associate/full professor | Economics | Register data, n = 17,467 academic economics | 1991–2008 |
| Carlsson et al. (2021) | Iceland, Norway, Sweden | Associate professor | Economics, Law, Physics, Political Science, Psychology, Sociology | Experimental data, 17 largest universities, country- and discipline-pooled analyses, n = 775 professors | 2018 |
| Filandri and Pasqua (2021) | Italy | Associate/full professor | All | Register data, mainly discipline-pooled analyses, n = 16,216 assistant professors and n = 3522 associate professors | Assistant/associate professors in 2012/13, followed until 2016 |
| Ginther and Kahn (2021) | US | Associate professor | Economics | Observational data, n = 798 assistant professors | Assistant professors in 2009, followed until 2018 |
| Henningsen et al. (2022) | Austria, Germany, Switzerland | Associate professor | Economics | Experimental data, country-pooled analyses, n = 481 from student assistant till full professor | 2020 |
| Lutter and Schröder (2016) | Germany | First tenured (associate/full) professorship | Sociology | CV data (webpage), pre-tenured sociologists, employed at German universities in 2013, n = 729 | 1980–2013 |
| Lutter et al. (2022) | Germany | First tenured (associate/full) professorship | Psychology | CV data (webpage), n = 2528 | 2018–2019 |

**Table A1** (*continued*)

| Authors | Country | Academic rank/ level | Disciplines | Data, method, sample size | Observation years |
|---|---|---|---|---|---|
| Moratti (2021) | Norway | Associate professor | "one faculty" | Observational data, one faculty at Norwegian University of Science and Technology, n = 1009 applicants | 2007–2017, all recruitments of new permanent associate/full professors by open call |
| Ooms et al. (2019) | Germany, Netherlands | (Assistant &) Associate and full professor | Disciplines present in the two universities considered | Observational survey data, one university per country, country- and discipline-pooled analyses, n = 248 academics (holding a PhD) | 2012–2013 |
| Perna (2001) | US | Associate/full professor | All | Observational data (1993 National Study of Postsecondary Faculty), discipline-pooled analyses, n = 8109 respondents | 1993 |
| Sarsons et al. (2021) | US | Associate professor | Economics, HRM | Observational CV data, n = 613 academic economics (35 departments), experimental data, n = 262 H R personnel | 1985–2014 |
| Schröder et al. (2021) | Germany | First tenured (associate/full) professorship | Political science | CV data (webpage), n = 1453 | 2018–2019 |
| Solga et al., 2023 | Germany | First full professorship | Mathematics, Physics, Economics, Social Sciences, German Studies | Experimental data, analyses by respondent's gender, mainly discipline-pooled analyses, n = 1688 professors | 2020 |
| Steinpreis et al. (1999) | US | (Assistant &) associate professor | Psychology | Experimental data, n = 238 professors | 1997 |
| Weisshaar (2017) | US | Associate professor | Sociology, Computer Science, English | Observational data (internet), n = 1560 assistant professors | Assistant professors in 2000–2004, contacted in 2014 |

**Table A2**

Correlation matrix of the vignette dimensions in the analytical samples

| Germany | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Applicants' gender | 1.000 | | | | | |
| 2. Publication record | 0.000 | 1.000 | | | | |
| 3. Parental leave | 0.004 | 0.005 | 1.000 | | | |
| 4. Collaborations | 0.004 | −0.005 | −0.002 | 1.000 | | |
| 5. Type of position | 0.000 | 0.000 | −0.004 | 0.003 | 1.000 | |
| 6. Research grant | −0.004 | 0.000 | −0.017 | −0.005 | 0.008 | 1.000 |

| Italy | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. Applicants' gender | 1.000 | | | | |
| 2. Publication record | −0.000 | 1.000 | | | |
| 3. Parental leave | 0.000 | 0.000 | 1.000 | | |
| 4. Collaborations | 0.000 | 0.000 | 0.009 | 1.000 | |
| 5. Type of position | 0.000 | 0.000 | −0.004 | −0.004 | 1.000 |

Note: Research grants in Italy are a fixed characteristic.

**Table A3**

Distributions of all vignette levels in the analytical samples

| Dimension | Levels | Germany | | Italy | |
|---|---|---|---|---|---|
| | | n | % | n | % |
| *Between-subject dimension* | | | | | |
| Applicants' gender | Male | 2052 | 50.0 | 648 | 51.2 |
| | Female | 2050 | 50.0 | 618 | 48.8 |
| *Within-subject dimensions* | | | | | |
| Publication record | Mostly single-authored | 1368 | 33.3 | 422 | 33.3 |
| | Mostly co-authored | 1364 | 33.2 | 422 | 33.3 |
| | All-co-authored | 1370 | 33.4 | 422 | 33.3 |
| Parental leave | Without parental leave (not reported in vignette) | 2043 | 49.8 | 633 | 50.0 |
| | With parental leave | 2059 | 50.2 | 633 | 50.0 |
| Research collaborations | With renowned scholars | 2059 | 50.2 | 633 | 50.0 |
| | With same-level scholars | 2043 | 49.8 | 633 | 50.0 |
| Type of position | With tenure-track. | 2050 | 50.0 | 633 | 50.0 |
| | Without tenure-track | 2052 | 50.0 | 633 | 50.0 |
| Research grant | Without a grant (not reported in vignette) | 2056 | 50.1 | | |
| | With a grant | 2046 | 49.9 | | |

Note: Only the German vignettes included the dimension "research grants." In Italy, grants (PhD fellowships) are very common at this stage (since most PhDs were funded with fellowships) and therefore fixed. In Germany, applicants are usually different in this respect Thus, the different treatment of grants in the German and Italian experiment contributes to the validity of the experiments and increases the comparability between countries. For the latter, we would argue that it reflects the real situation in both countries and thus "harmonizes" what professors were asked to evaluate.

**Table A4a**
Invitation letters

| Invitation letter – Italy |
| --- |
| Hello, |
| We are a group of researchers from the European University Institute, the University of Lausanne, and the WZB Berlin Social Science Center conducting a study on academic culture in different international contexts. |
| For the research to be successful, we need high participation in the survey, and given your position in the research and academic field, we believe your participation could provide valuable insights. |
| The survey will be conducted using the Qualtrics® software. Participation is voluntary, and the information gathered will be kept strictly confidential. |
| Completing the questionnaire will only take 10 min of your time. |
| Thank you in advance for your commitment and cooperation. As a token of our appreciation for your assistance, we would be pleased to offer you the chance to participate in a drawing for an iPad Pro 11″ Wi-Fi 256 GB. The results of the drawing will be announced upon completion of data collection. |
| All information is collected in accordance with current regulations on the protection of personal data and will only be used for statistical purposes. |
| To participate, please follow the link below: Take the Survey |
| Best regards, |

| Invitation letter – Germany |
| --- |
| Dear colleague Prof. [last name], |
| We would like to invite you to participate in a survey on selection criteria in appointment procedures for professorships. In light of the current situation, we would also like to find out how the COVID 19 pandemic is affecting your everyday working life. |
| The survey is conducted by the Berlin Social Science Center (WZB) and the German Center for Higher Education Research and Science Studies (DZHW). You are invited to participate in this survey because you have an internet presence at a German university, or had one in the summer semester of 2019. |
| Your participation is very important to us because, as a professor, you have important expertise on professorship appointment procedures and standards. |
| The following link will direct you to the online questionnaire (completion time approximately 20 min): [link] |
| Please do not share this link. It is personalized and allows you to participate only once. |
| Your data will of course be kept completely confidential and anonymous. Further information about data protection and consent can be found at the beginning of the survey. |
| Should you have any questions or feedback, please do not hesitate to contact us (email). |
| Thank you very much for your participation! |
| Best regards, [names] |

**Table A4b**
Vignette examples

| *Before the vignette block:* |
| --- |
| Suppose you are a member of an appointment committee for a new assistant professorship in your faculty or department. We will now show you short descriptions of some applicants for an assistant professorship in your discipline. |
| Please read each profile carefully and evaluate each applicant individually. In total, we will present 6 applicants. |
| *Vignette:* |
| **Italy** |
| It is an assistant professorship <u>with tenure track</u>. |
| Giorgia Bianchi has a PhD from an Italian university and 6 years of postdoc experience. She has obtained national grants for conducting her research, which has resulted in 10 peer-reviewed articles: <u>7 single authored and 3 co-authored</u>. Moreover, she has established <u>research collaborations with scholars of similar academic level</u> and has accumulated teaching experience at the Bachelor and Master levels. |
| **Germany** |
| It is an assistant professorship <u>without tenure track</u>. |
| Lea Fuchs received a PhD (graded) magna cum laude from a German university. Her CV shows 4 years of postdoc experience <u>and 6 months of parental leave</u>. Her research is published in {number} peer-reviewed articles, of <u>which 70% are single-authored and 30% are co-authored</u>. The applicant has <u>research collaborations with renowned scholars</u> and teaching experience at Bachelor and Master levels. <u>She has successfully obtained third-party funding</u>. |
| *After each vignette:* |

**Table A4b** (*continued*)

| *Before the vignette block:* |
| --- |

How would you rate this applicant on a scale of 1–7, with 1 being the most negative and 7 being the most positive score?

(1) Would you invite this applicant for an interview? ⟨< 1–7 ≫

(2) Do you think this applicant is qualified for the position? ⟨< 1–7 ≫

Note: Underlined are the vignette dimensions in these examples. In the surveys, they were not underlined. {number} = respondent-specific number (for explanation, see Data and Method section).

# References

Acker, Joan, 2006. "Inequality Regimes: Gender, Class, and Race in Organizations." Gender & Society 20 (4), 441–464.

Aksnes, Dag W., Kahn, Shulamit, Borgan Reiling, Rune, Marte, E., Ulvestad, S., 2022. "Longitudinal evidence on Norwegian PhDs suggests slower progression for women academics but not a leaky pipeline.". https://doi.org/10.31235/osf.io/pvx8q. (Accessed 4 October 2022).

Arrow, Kenneth J., 1973. In: Ashenfelter, O., Rees, A., Princeton, N.J. (Eds.), "The Theory of Discrimination." Pp. 3–33 in Discrimination in Labor Markets. Princeton University Press.

Auspurg, Katrin, Hinz, Thomas, 2015. Factorial Survey Experiments. SAGE, Los Angeles.

Auspurg, Katrin, Hinz, Thomas, Schneck, Andreas, 2017. "Berufungsverfahren als turniere.". Z. Soziol. 46 (4), 283–302.

Bosquet, Clément, Combes, Pierre-Philippe, García-Peñalosa, Cecilia, 2019. "Gender and Promotions: Evidence from Acad. Econ. France." Scandinavian J. Econ. 121, 1020–1053.

Bosson, Jennifer K., et al., 2021. "Psychometric properties and correlates of precarious manhood beliefs in 62 nations.". Journal of Cross-Cultural Psychology 52 (3), 231–258.

Brink, Van den, Marieke, Benschop, Yvonne, Jansen, Willy, 2010. "Transparency in academic recruitment.". Organ. Stud. 31 (11), 1459–1483.

Bundesamt, Statistisches, 2018. Personal an Hochschulen 2017. Statistisches Bundesamt, Wiesbaden.

Card, D., DellaVigna, S., Funk, P., Iriberri, N., 2020. Are referees and editors in economics. J. Econ. 135 (1), 269–327.

Card, D., DellaVigna, S., Funk, P., Iriberri, N., 2023. Gender gaps at the academies. Proc. Natl. Acad. Sci. 120 (4), e2212421120.

Carey, John M., Carman, Kevin R., Clayton, Katherine P., Horiuchi, Yusaku, Htun, Mala, Ortiz, Brittany, 2020. "Who wants to hire a more diverse faculty?". Politics, Groups, and Identities 8 (3), 535–553.

Carlsson, Magnus, Henning, Finseraas, Midtbøen, Arnfinn H., Rafnsdóttir, Guðbjörg L., 2021. "Gender bias in academic recruitment? Evidence from a Survey Experiment in the Nordic Region." European Sociol. Review 37 (3), 399–410.

Ceci, Stephen J., 2018. "Women in Acad. Sci.: Experimental Findings from Hiring Stud." Educ. Psychol. 53 (1), 22–41.

Ceci, Stephen J., Williams, Wendy M., 2015. "Women have substantial advantage in STEM faculty hiring. Except When Competing Against More-Accomplished Men." Front. Psychol. 6, 1532.

Ceci, Stephen J., Kahn, Shulamit, Williams, Wendy M., 2023. "Exploring gender bias in six key domains of academic science.". Psychol. Sci. Publ. Interest 24 (1), 15–73.

Chang, Edward H., et al., 2019. "Diversity thresholds: how social norms, visibility, and scrutiny relate to group composition.". Acad. Manag. J. 62 (1), 144–171.

Clogg, Clifford C., Petkova, Eva, Haritou, Adamantios, 1995. "Statistical methods for comparing regression coefficients between models.". Am. J. Sociol. 100, 1261–1293.

Correll, Shelley J., Benard, Stephen, 2006. "Biased estimators? Comparing status and statistical theories of gender discrimination.". In: Thye, S.R., Lawler, E.J. (Eds.), Advances in Group Processes,, vol. 23. Elsevier JAI, Amsterdam, pp. 89–116.

Correll, Shelley J., Ridgeway, Cecilia L., 2003. In: Delamater, J. (Ed.), "Expectation States Theory." Pp. 29–51 in Handbook Of Social Psychology. Springer, Boston.

Correll, Shelley J., Benard, Stephen, In Paik, 2007. "Getting a Job: Is There a Motherhood Penalty?". Am. J. Sociol. 112 (5), 1297–1338.

Crandall, Christian S., Eshleman, Amy, 2003. "A justification-suppression model of the expression and experience of prejudice.". Psychol. Bull. 129 (3), 414–446.

Crowley, Niall, Sansonetti, Silvia, 2019. In: New Visions for Gender Equality 2019. European Union, Brussels.

Cuddy, Amy J., Susan, T. Fiske, Glick, Peter, 2004. "When professionals become mothers, Warmth Doesn't Cut the Ice.". J. Soc. Iss. 60 (4), 701–718.

European Commission, 2014. Gender equality policies in public research. Luxemburg: Publications Office of the European Union.

European Commission, 2016. She Figures 2015. Brussels. Publications Office of the European Union.

European Commission, 2019. She Figures 2018. Brussels. Publications Office of the European Union.

European Commission, 2020. A Union of equality: gender equality strategy 2020-2025. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX: 52020DC0152&from=EN. (Accessed 2 February 2023).

European Commission, 2021. She Figures 2021. Brussels. Publications Office of the European Union.

Farré, Lídia, Ortega, Francesc, 2023. "Selecting talent. Gender differences in success in competitive selection processes.". J. Hum. Resour. 58 (6), 1881–1913.

Filandri, Marianna, Pasqua, Silvia, 2021. "'Being Good Isn't Good Enough': Gender Discrimination in Italian Academia.". Studies in Higher Education 46 (8), 1533–1551.

Foschi, Martha, 1996. "Double standards in the evaluation of men and women.". Soc. Psychol. Q. 59 (3), 237–254.

Foschi, Martha, Lai, Larissa, Sigerson, Kirsten, 1994. "Gender and double standards in the assessment of job applicants.". Soc. Psychol. Q. 57 (4), 326–339.

Gerxhani, Klarita, Kulic, Nevena, Liechti, Fabienne, 2023. "Double standards? Co-authorship and gender bias in early-stage academic evaluations.". Eur. Socio Rev. 39 (2), 194–209.

Ginther, Donna K., Kahn, Shulamit, 2021. "Women in academic economics". AEA Pap. Proc. 111, 138–142.

Glazer, Judith S., 2005. In: Marshall, C. (Ed.), "Affirmative Action and the Status of Women in the academy." Pp. 67–80 in Feminist Critical Policy Analysis II. Routledge, London.

GWK, 2020. Chancengleichheit in Wissenschaft und Forschung (Gemeinsame Wissenschaftskonferenz). Materialien der GWK. GWK, Heft 69. Bonn.

Henningsen, Levke, Horvath, Lisa K., Jonas, Klaus, 2022. "Affirmative action policies in academic job advertisements.". Sex. Roles 86 (1–2), 34–48.

Herschberg, Channah, Benschop, Yvonne, Brink, Marieke van den, 2018. "Precarious postdocs: a comparative study on recruitment and selection of early-career researchers.". Scand. J. Manag. 34 (4), 303–310.

Hipp, Lena, 2018. "Damned if you do, damned if you don't? Experimental evidence on hiring discrimination against parents with differing lengths of family leave.". https://doi.org/10.31235/osf.io/qsm4x. (Accessed 27 August 2018).

Koch, Amanda J., Susan, D., Mello, D., Sackett, Paul R., 2015. "A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making.". J. Appl. Psychol. 100 (1), 128–161.

Kosakowska-Berezecka, Natasza, et al., 2023. "Gendered self-views across 62 countries: a test of competing models.". Soc. Psychol. Personality Sci. 14 (7), 808–824.

LERU, 2018. Implicit bias in academia: a challenge to the meritocratic principle and to women's careers. Advice paper no. 23. Leuven: league of European research universities. https://www.leru.org/files/Publications/Implicit-bias-in-academia-Full-Paper.pdf. (Accessed 2 December 2022).

Leslie, Lisa M., Flaherty Manchester, Colleen, Dahm, Patricia C., 2017. "Why and when does the gender gap reverse? Diversity goals and the pay premium for high potential women." Academy of Management Journal, 60.2, 402–432.

Lutter, Mark, Schröder, Martin, 2016. "Who becomes a tenured professor, and why? Panel data evidence from German sociology, 1980–2013.". Res. Pol. 45 (5), 999–1013.

Lutter, Mark, Habicht, Isabel M., Schröder, Martin, 2022. "Gender differences in the determinants of becoming a professor in Germany. An Event History Analysis of Academic Psychologists from 1980 to 2019.". Research Policy 51 (6), 104506.

Marini, Giulio, Meschitti, Viviana, 2018. "The Trench Warfare of Gender Discrimination: Evidence from Academic Promotions to Full Professor in Italy.". Scientometrics 115 (2), 989–1006.

Merton, Robert K., 1995. "The Thomas theorem and the matthew effect.". Soc. Forces 74 (2), 379–422.

Morana, Marie T., Sagramora, Simonetta, 2020. Focus 'le carriere femminili in ambito accademico'. Ministero dell'Istruzione, dell'Università e della Ricerca. Gestione patrimonio informativo e statistica, Italy.

Moratti, S., 2021. A woman's place is in the 'home'? Gender-specific hiring patterns in academia in gender-equal Norway. J. Sociol. 57 (4), 916–934.

Morgan, Allison C., Way, Samuel F., Hoefer, Michael J.D., Larremore, Daniel B., Galesic, Mirta, Clauset, Aaron, 2021. "The unequal impact of parenthood in academia.". Sci. Adv. 7 (9) eabd1996.

Murgia, Annalisa, Poggio, Barbara, 2019. Gender and Precarious Research Careers. Routledge, London.

NASEM, 2020. National academies of sciences, engineering, and medicine. Promising Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine: Opening Doors. The National Academies Press, Washington, DC.

OECD, 2022. In: OECD Family Database: PF2.1. Parental Leave Systems. OECD, Paris. https://www.oecd.org/els/soc/PF2_1_Parental_leave_systems.pdf. (Accessed 8 April 2023).

Ooms, Ward, Werker, Claudia, Hopp, Christian, 2019. "Moving up the ladder.". Stud. High Educ. 44 (7), 1268–1289.

Orupabo, Julia, Mangset, Marte, 2022. "Promoting Diversity but Striving for Excellence: Opening the 'Black Box' of Academic Hiring.". Sociology 56 (2), 316–332.

Paternoster, Raymond, Brame, Robert, Mazerolle, Paul, Piquero, Alex, 1998. "Using the correct statistical test for the equality of regression coefficients.". Criminology 36, 859–866.

Perna, Laura W., 2001. "Sex and race differences in faculty tenure and promotion.". Res. High. Educ. 42 (5), 541–567.

Phelps, Edmund S., 1972. "The statistical theory of racism and sexism.". Am. Econ. Rev. 62, 659–661.

Ridgeway, Cecilia L., 2019. Status: Why Is it Everywhere? Why Does it Matter? New York: Russell Sage Foundation.

Ridgeway, Cecilia L., Correll, Shelley J., 2004. "Motherhood as a status characteristic.". J. Soc. Issues 60, 683–700.

Ridgeway, Cecilia L., Markus, Hazel Rose, 2022. "The significance of status.". RSF: The Russell Sage Foundation Journal of the Social Sciences 8 (7), 1–25.

Ross, Matthew B., Glennon, Britta M., Murciano-Goroff, Raviv, Berkes, Enrico G., Weinberg, Bruce A., Lane, Julia I., 2022. "Women Are Credited Less in Science Than Men.". Nature 608, 135–145.

Sarsons, Heather, 2017. "Recognition for Group Work: Gender Differences in Academia.". Am. Econ. Rev. 107 (5), 141–145.

Sarsons, Heather, Gërxhani, Klarita, Reuben, Ernesto, Schram, Arthur, 2021. "Gender differences in recognition for group work.". J. Polit. Econ. 129 (1), 101–147.

Schaerer, Michael, du Plessis, Christilene, Nguyen, My Hoang Bao, Robbie, C.M., van Aert, Leo, Tiokhin, Elena Giulia Clemente, Daniël Lakens, Pfeiffer, Thomas, Magnus Johannesson, Anna Dreber, Clark, Cory J., Uhlmann, Eric Luis, 2023. "On the Trajectory of Discrimination: A Meta-Analysis and Forecasting Survey Capturing 44 Years of Field Experiments on Gender and Hiring Decisions.". Org. Behav. Hum. Decis. Proc. 179, 104280.

Schröder, Martin, Lutter, Mark, Habicht, Isabel M., 2021. Determinants of becoming a tenured professor in German political science.". PLoS One 16 (1), e0243514. "Publishing, Signaling, Social Capital, and Gender.

Silva, Virgínia, Santos, Maria, Rosa, Miriam, 2021. "Does she deserve it? The Influence of Gender and Meritocracy in Reactions to Affirmative Action Legislation.". Soc. Sci. 10 (10), 386.

Solga, Heike, Rusconi, Alessandra, Netz, Nicolai, 2023. Professors' gender biases in assessing applicants for professorships. Eur. Sociol. Rev. 39 (6), 841–861.

Steinpreis, Rhea E., Anders, Kati A., Ritzke, Dawn, 1999. "The impact of gender on the review of the curricula vitae of job applicants and tenure candidates.". Sex. Roles 41, 509–528.

Su, Dan, Steiner, Peter M., 2020. "An evaluation of experimental designs for constructing vignette sets in factorial surveys.". Socio. Methods Res. 49 (2), 455–497.

Treischl, Edgar, Wolbring, Tobias, 2022. "The past, present and future of factorial survey experiments.". Methods, data, analyses 16 (2), 141–170.

Urry, Meg, 2015. "Science and gender: scientists must work harder on equality.". Nature 528 (7583), 471–473.

Walzenbach, Sandra, 2019. "Hiding sensitive topics by design? An Experiment on the Reduction of Social Desirability Bias in Factorial Surveys.". Survey Research Methods 13 (1), 103–121.

Weisshaar, Katherine R., 2017. "Publish and perish? An assessment of gender gaps in promotion to tenure in academia.". Soc. Forces 96 (2), 529–560.

Weisshaar, Katherine R., 2018. "From opt out to blocked out: the challenges for labor market Re-entry after family-related employment lapses.". Am. Socio. Rev. 83 (1), 34–83.

Williams, Wendy M., Ceci, Stephen J., 2015. "STEM faculty prefer hiring women professors 2.1.". Proc. Nat. Acad. Sci. 112 (17), 5360–5365.