

Solga, Heike; Rusconi, Alessandra; Hofmeister, Sophie

**Article — Published Version**

## Gender biases in assistant professor recruitment: Does discipline matter?

Research Policy

**Provided in Cooperation with:**

WZB Berlin Social Science Center

*Suggested Citation:* Solga, Heike; Rusconi, Alessandra; Hofmeister, Sophie (2025) : Gender biases in assistant professor recruitment: Does discipline matter?, Research Policy, ISSN 1873-7625, Elsevier, Amsterdam, Vol. 54, Iss. 3, pp. 1-17,  
<https://doi.org/10.1016/j.respol.2024.105170>

This Version is available at:

<https://hdl.handle.net/10419/315853>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



# Gender biases in assistant professor recruitment: Does discipline matter?

Heike Solga<sup>a,b,\*</sup>, Alessandra Rusconi<sup>b</sup>, Sophie Hofmeister<sup>b</sup>

<sup>a</sup> Freie Universität Berlin, Institute of Sociology, Germany

<sup>b</sup> WZB – Berlin Social Science Center, Department “Skill Formation and Labor Markets”, Germany

## ARTICLE INFO

### Keywords:

Gender bias  
Assistant professorships  
Disciplinary comparison  
Prejudice expression model  
German academia

## ABSTRACT

Higher education institutions have implemented various affirmative action policies aimed at increasing the representation of female professors, including measures to reduce gender bias in professorship appointments. This raises the question of whether gender bias still exists. Research on gender bias in assistant professor appointments remains sparse. We therefore examine whether gender bias in assistant professor recruitment exists and differs across disciplines (looking at mathematics/physics, economics/sociology/political science, and German studies). Our analysis is based on a factorial survey experiment with 1857 professors from German universities in 2020. We draw on Crandall and Eshleman's (2003) justification-suppression model to argue that gender policies can help suppress the expression of prejudices (negative stereotypes) against female applicants. Our results show that in all disciplines studied, female applicants receive higher ratings than male applicants, both for perceived qualification for an assistant professorship and for being invited for an interview. The female advantage is more pronounced in mathematics/physics when applicants are perceived to be equally qualified, suggesting a greater normative pressure to comply with gender-based preferential selection. In mathematics/physics, however, we also find a smaller premium for having received a research grant among female applicants. Overall, the observed female advantage is rather small in all disciplines studied.

## 1. Introduction

The underrepresentation of women among professors is a widely recognized problem in terms of productivity, diversity, and equity in academia (European Commission, 2020; LERU, 2018). For several years, higher education institutions (HEIs) have implemented numerous affirmative action measures to increase the proportion of female professors, including measures to prevent gender bias in professorship appointments (European Commission, 2014; European Commission, 2021). However, the underrepresentation of female professors remains considerable—only 26 % of full professors in the EU are women (in 2018). The underrepresentation varies considerably across academic disciplines: it is highest in engineering (17 %) and natural sciences (21 %), and lowest, but still noticeable, in the humanities (34 %) (European Commission, 2021, pp. 184, 190). Both phenomena raise the question of whether gender bias still exists in professorship appointments (i.e.,

whether women are less likely to be hired even when they apply for professorships) and whether this differs across disciplines. Or, alternatively, whether women's lower propensity to apply for professorships is the main reason for their underrepresentation and the disciplinary differences.

The prevailing narrative in this debate claims that (un)conscious gender bias disadvantages female applicants in professorship appointments (LERU, 2018; National Academies of Sciences, Engineering, and Medicine, 2020). Contrary to this narrative, a growing body of research suggests that the primary cause of this underrepresentation is not gender bias in appointments; rather, women are less likely than men to apply for professorships (see (Ceci et al., 2021; Ceci et al., 2023)). More and more studies even document a female advantage in faculty recruitment. Some of these studies are from Europe, with more recent data and experimental designs. Examples include several observational studies of tenured faculty appointments in Germany for political science,

\* Corresponding author at: WZB, Reichpietschufer 50, 10785 Berlin, Germany.

E-mail address: [heike.solga@wzb.eu](mailto:heike.solga@wzb.eu) (H. Solga).

<https://doi.org/10.1016/j.respol.2024.105170>

Received 16 February 2023; Received in revised form 20 December 2024; Accepted 22 December 2024

Available online 15 January 2025

0048-7333/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

psychology, and sociology (Habicht et al., 2024; Lutter et al., 2022; Schröder et al., 2021) and across all disciplines at one university (Auspurg et al., 2017), and three factorial survey experiments (also called vignette studies) for associate or full professor appointments, that either pooled different disciplines or across countries, or both (Carlsson et al., 2021; Henningsen et al., 2022; Solga et al., 2023).

Compared to (tenured) faculty hiring, research on gender bias in assistant professor appointments is very scarce. Given the bottlenecks in the academic system, discriminatory recruitment processes at this transition stage may contribute to the leaky pipeline at later stages, as assistant professors represent an important applicant pool for future associate/full professors (Williams and Ceci, 2015). While assistant professorships are not yet tenured positions in many countries, the recruitment process follows the same formalized hiring procedures as for tenured faculty and should similarly comply with HEI's gender equality policies (Herschberg et al., 2018).

Again, and despite this lack of research, the prevailing narrative is that “there is an undeniable body of evidence showing that bias against women operates in recruitment and selection processes already for early-career female researchers.” (LERU, 2018, p. 4) However, only two studies find a male advantage (i.e., discrimination against women): the older US study for tenure-track assistant professorships in psychology by Steinpreis et al. (1999) and the study by Ooms et al. (2019) based on pooled analyses of 2012/13 survey data from one German and one Dutch technical university. In contrast to this discrimination narrative, several studies even find a female advantage. US examples are the conjoint survey experimental study by Carey et al. (2020) conducted at two public universities in 2016/17, and several experimental studies for STEM disciplines conducted in 2011/12 by Williams and Ceci (2015) (see also Ceci, 2018; Ceci and Williams, 2015). A recent factorial survey experiment for assistant professor recruitment in the social sciences also found a female advantage in Germany, and no gender difference in Italy (Görxhani et al., 2025).

The results of these studies may differ because they focus on different disciplines—for example, male- vs. female-dominated disciplines—on the average across different disciplines. Existing research rarely compares across disciplines or has a limited range, like Williams and Ceci (2015), who compared math-intensive and non-math-intensive fields (i. e., engineering/economics vs. psychology/biology). Comparisons across studies are problematic due to differences in methods used or countries included. Disciplinary comparisons, however, are particularly important because, at least for the United States, there is some evidence that “it is the very fields in which women are well-represented where there are hints of bias in the sense that more earn PhDs than are hired as assistant professors.” (Ceci et al., 2021, p. 40) It is unclear, however, whether this is due to gender bias in hiring decisions or a lower likelihood of women applying for jobs in these disciplines.

Against this background, we examine *whether there is a female or male advantage in assistant professor recruitment and whether this differs across disciplines*. We use Crandall and Eshleman's (2003) justification-suppression model of the expression of prejudice as our theoretical foundation for examining differences across disciplines. Following this theory, we do not argue that professors from different disciplines hold more or less strong gender-stereotypical prejudices, or that gender policies change gatekeepers' prejudices against women. We merely assume that gender policies contribute to suppressing the expression of prejudices against female applicants and thus their impact on appointment processes, and that the degree of suppression may differ depending on the disciplinary context.

Our analyses are based on data from a factorial survey experiment (hereafter “factorial survey”) that we conducted in 2020 with 1857 professors from very different disciplines in Germany. We selected disciplines that differ in their share of female researchers, their publication styles with respect to co-authorship, and come from different disciplinary areas: mathematics and physics (from the natural sciences), economics, political science, and sociology (from the social sciences),

and German studies (from the humanities). Similar to Carlsson et al. (2021) and Henningsen et al. (2022), we asked the professors to evaluate descriptions of hypothetical applicants for assistant professorships<sup>1</sup> on two aspects: “perceived qualification” (i.e., whether they consider the applicant to be qualified for an assistant professorship) and “invitation propensity” (i.e., how likely they would invite the applicant to a job interview). Like Carlsson et al. (2021), we combine a between-subject design for the applicant's gender and a within-subject design for all other profile dimensions to avoid the risk of socially desirable (gender) responses.

The paper is organized as follows: We begin with some insights into the German academic system, along with official statistics on gender differences in assistant professorships across disciplines. We then develop theoretical ideas and hypotheses about why the applicants' gender may influence evaluators' perceptions of applicants' qualification and their propensity to invite them for an interview differently across disciplines. In the next two sections, we describe our research design, sample, and analytical strategy, and then present the results. We conclude with a summary, our contributions to the existing literature, and some directions for future research, while also reflecting on the limitations of our study.

## 2. The German context

Germany represents a country with a rather strong normative climate for gender equality in general and in academia. Our descriptive analysis of the 2017 European Values Study<sup>2</sup> (only including respondents with a tertiary degree, to be as close as possible to the educational level of our respondents) reveals widely shared gender equality norms in Germany: Only 19 % of German respondents (strongly) agreed with the statement “When a mother works for pay, the children suffer,” and 90 % (strongly) disagreed with the statement “When jobs are scarce, men have more right to a job than women.”

This normative climate is mirrored by policies aimed directly at advancing women in academia (Henningsen et al., 2022): 89 % of German HEIs have implemented affirmative action measures/policies, compared to, for example, only 52 % of Dutch HEIs or 69 % of French HEIs (European Commission, 2021, p. 170). These policies include a law first implemented in North Rhine-Westphalia in 1989 and meanwhile implemented at the federal level as § 42 of the Framework Act for Higher Education (*Hochschulrahmengesetz*). The law requires HEIs to prefer women in cases of equal performance of male and female applicants in areas where they are underrepresented—also referred to as “weak gender-based preferential selection” (in contrast to “strong preferential selection,” which applies to unequally qualified female candidates; Silva et al., 2021). In addition, there are gender-sensitive appointment guidelines; equal opportunity officers who participate in the universities' appointment committees; regular monitoring within HEIs and, since 1989, by the Joint Science Conference (*Gemeinsame Wissenschaftskonferenz*); and the provision of implicit bias training for researchers/professors. In 2008, the German Research Foundation (DFG) (the most important research funding organization) introduced “Research-Oriented Gender Equality Standards” with the ultimate goal of increasing women's representation in professorships, and all public universities have committed themselves to these standards.

Despite these policies, women are underrepresented among professors, as in other academic systems. In 2020, the female share among full professors at German universities was only 22 % on

<sup>1</sup> There are three different ranks of professorships in Germany: W1, W2, and W3, corresponding to assistant, associate, and full professorships in the international nomenclature. This study focuses on W1 professorships.

<sup>2</sup> 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>.

average—ranging from 35 % in the humanities, 26 % in the social sciences (including economics and law), 17 % in mathematics/natural sciences, to only 11 % in engineering (Statistisches Bundesamt, 2020, pp. 46, 50). Among assistant professors, the proportion of women was close to parity (48 %), but again it varied widely across disciplines—from 64 % in the humanities, 48 % in the social sciences, 40 % in mathematics/natural sciences, to only 27 % in engineering.

These overall figures for assistant professorships may mask important gender differences when it comes to tenure. The German academic system is classified as a “competition model,” meaning that the number of associate and full professorships is limited/fixed (Aksnes et al., 2022, p. 3). All assistant professors are hired on a fixed-term contract, with one important distinction—namely, whether their position includes a tenure track. For a tenured (associate/full) professorship, non-tenure-track assistant professors have to apply to job openings that depend on the availability of “vacant” positions at universities. In contrast, the promotion of tenure-track assistant professors to an associate or even full professorship depends only on the (evaluation of the) quality of their research record. This tenure-track/non-tenure-track distinction may be crucial for gender inequality in career progression.

In Germany, only 29 % of assistant professorships included a tenure track in 2020 (Statistisches Bundesamt, 2020, p. 46). Official statistics show not only considerable differences between disciplines—with 28 % tenure-track assistant professors in the social sciences and 29 % in the humanities, but 36 % in mathematics/natural sciences—but also between female and male assistant professors. In the humanities, only 24 % of female assistant professors were tenure-track, compared to 37 % of male assistant professors, while the percentages of female and male

tenure-track assistant professors were very similar in the social sciences (29 % vs. 28 %) and mathematics/natural sciences (38 % vs. 36 %). We have therefore included the type of assistant professorship as a dimension in our study (see Section 4.2).

All statistics presented so far reflect the combined outcomes of application and recruitment behavior. Thus, they do not indicate whether gender bias exists in evaluations of applicants for assistant professorships and whether this differs across disciplines. Fig. 1, which presents recruitment statistics rather than stock statistics (GWK, 2019; GWK, 2021), seems to indicate that female applicants are favored in being hired as assistant professors (conditional on having applied): The proportion of women increases with each phase of the appointment process. On average across all disciplines, the proportion of women among the “offers” for assistant professorships (47.7 %) is almost 50 % higher than the proportion of women among the “applications” (32.8 %). This strong increase is observable not only in the humanities and social sciences, but even more so in mathematics/natural sciences.

Fig. 1 also reveals that the proportion of applications from women is lower than their proportion among PhD recipients in 2016 and 2017 (a reasonable time before transitioning to assistant professorships in 2019/2020). This gap is particularly large in mathematics/natural sciences, still considerable in the social sciences, but much smaller in the humanities. This seems to strengthen the idea that disciplinary differences in women’s propensity to apply, rather than discrimination against female applicants during the recruitment process, are today the main source of differences in the proportion of female professors across disciplines.

Fig. 1, however, does not tell us whether the observed female advantage actually indicates that there is no discrimination against

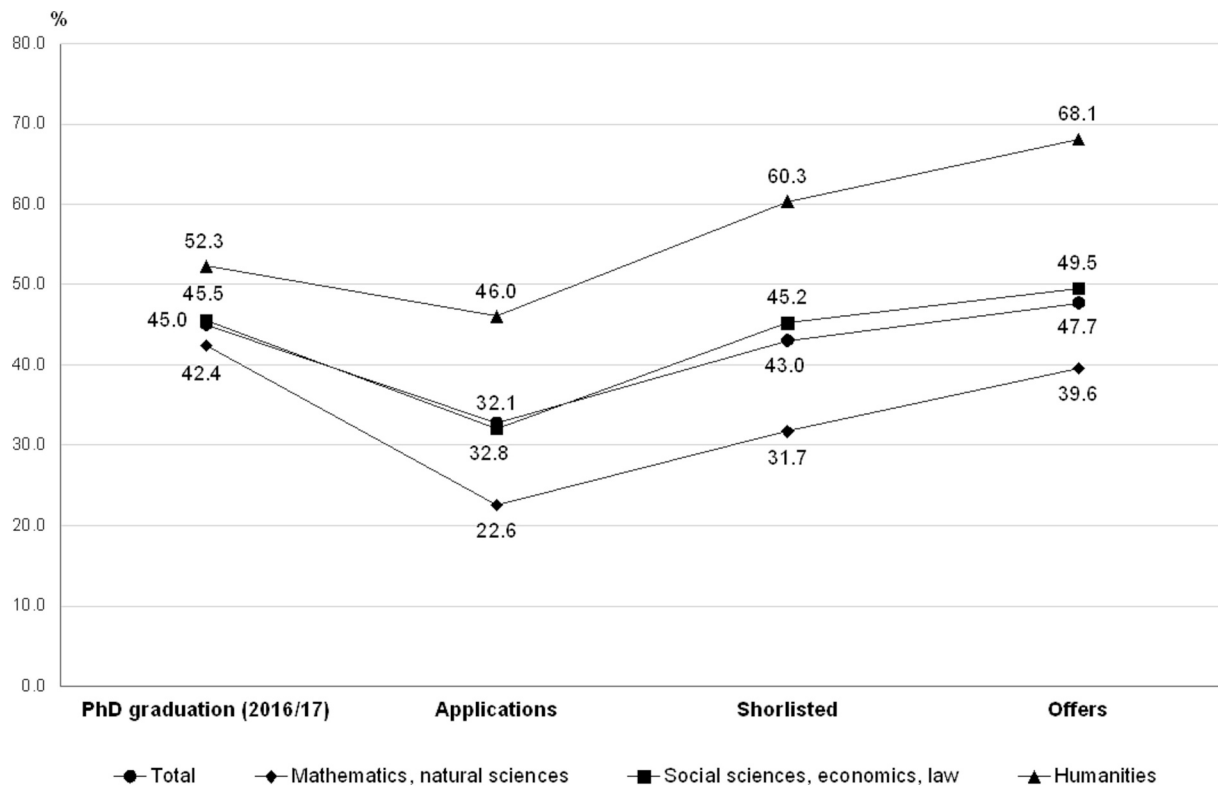


Fig. 1. Proportion of women throughout recruitment phases for assistant professorships in Germany, 2019/20 (in %). Source: GWK (2019, 2021).

women or even a preferential treatment of women, because the units of analysis are applications and not individuals. The female advantage could be due to men applying more often than women, for example, because men are less selective about their suitability for positions (Urry, 2015, p. 437). Moreover, the cohort of female applicants, while smaller, may be qualitatively stronger than their male competitors. Such a “survivor bias” (Auspurg et al., 2017, p. 284) may be more pronounced in mathematics/natural sciences, given the larger difference between the proportion of PhDs and of assistant professorship applications, and may explain the steeper increase in the proportion of women from applications to offers here. The higher proportion of women who are ultimately offered assistant professorships (compared to their proportion among the applicants) masks the possibility that few women receive multiple offers from different universities, and they can, of course, accept only one. Thus, it remains unclear whether there is a gender bias in recruitment processes between *equally* qualified male and female applicants and whether this differs across disciplines—which is what our study aims to investigate.

To answer these questions, we need to have identical application behavior and identical performance for men and women, i.e., identical applicant pools (Koch et al., 2015). We therefore used a factorial survey, which allows us to keep the academic profiles of male and female applicants equal and thus to simulate identical application behavior for both. Moreover, it allows us to keep the applicant pool identical across disciplines. Thus, any gender differences found in the analyses are then due to gender bias, and any differences between disciplines are due to differences in the evaluation of male and female applicants (and not due to differences in how many women apply and/or differences in applicants’ quality). Furthermore, this experimental design is consistent with the core idea of the justification-suppression model of prejudice expression discussed in the next section, since we observe the expression of prejudice in respondents’ evaluative behavior rather than prejudice *per se*.

### 3. Theoretical considerations and hypotheses

Several theories argue for the existence of (descriptive and prescriptive) gender-science stereotypes that ultimately lead to biases against women in the evaluation of scholarly performance. The stereotype content model (Fiske et al., 2002) and the lack of fit model (Heilman, 2012) identify culturally shared gender stereotypes as a major reason why men are (still) more often associated with agentic character traits such as being rational, competent, efficient, and goal-oriented than women. Women, by contrast, are (still) more often associated with communal traits such as being empathic, emotionally sensitive, friendly, and deferential. Studies have shown that leadership and professorship positions are typically associated with agentic traits, that is, the same traits attributed to men (Carli et al., 2016; Miller et al., 2015; Smyth and Nosek, 2015), and that these traits are considered necessary to meet the requirements of higher-level positions (Fiske and Dupree, 2014). Expectation states theory (Berger et al., 1972; Correll and Ridgeway, 2003) adds that even when information about applicants’ performance is fully available, applicants’ gender and associated stereotypical attributions can serve as a strong proxy for academic performance. This may lead to biased evaluations of women through the use of “double standards” in evaluating women and men, meaning that stricter standards are applied to women’s performance even when complete information about both women’s and men’s performance is available (Foschi, 2000). Following these theories, we generally assume that gender-science stereotypes are prevalent among academic gatekeepers.

Gender prejudices refer to the negative aspects of gender stereotypes. Marx and Ko (2019) argue that prejudices often remain unchanged even in the face of contradictory information, for example, because they simplify the interpretation of individuals’ behavior based on their group membership. However, according to Crandall and Eshleman’s (2003) justification-suppression model of prejudice expression, “commitment to nonexpression of prejudice (...) may not be congruent with stereotypes”—because prejudices are “fluid” in the sense that situational contexts may influence whether individuals express or endorse prejudices (Marx and Ko, 2019, p. 1). The core idea of this model is that people are motivated to suppress prejudices when the prevailing social norms in a person’s environment are negative toward those prejudices—simply expressed by the equation: “prejudice + suppression = expression” (Crandall and Eshleman, 2003, p. 416). Prejudice suppression should not be confused with the behavior of social desirability in surveys. Social desirability refers to the intentional concealment of true beliefs and opinions that one would otherwise have expressed. In contrast, prejudice suppression is an individual’s attempt to “restrain admission of prejudice into the self-concept” and thus to banish the appearance of prejudice from the self (Crandall and Eshleman, 2003, p. 421).

In applying the justification-suppression model to our research question, we do not assume that the strength of prejudices against women varies across disciplines (i.e., we do not assume that professors—the members of appointment committees—choose their disciplines because of their gender stereotypes). However, differences may exist across disciplines in the extent to which professors suppress *expressions* of gender-science prejudices in their evaluations (e.g., because of disciplinary differences in the gender distribution of professors, see below).

Across disciplines, it could be argued that affirmative action policies can foster a more gender-equality-friendly normative climate at HEIs (Crosby et al., 2006; Henningsen et al., 2022) and among professors (Koch et al., 2015; Williams and Ceci, 2015). Gender policies as expressions of organizational values and a gender-equality normative climate may not change gatekeepers’ prejudices against female applicants, but they may reduce stereotyping and prejudice expression (Koch et al., 2015). A normative climate in which expressing gender-science stereotypes/prejudices is considered old-fashioned and/or socially unacceptable may externally/internally motivate individuals to suppress gender prejudices when evaluating applicants. And this, in turn, may be the reason why gender policies reduce gender bias in professorial appointments. Based on these considerations, we expect to find the following for *all disciplines* included in our study, indicating that suppression of gender-science prejudice expression exists in German academia:

- a female advantage in the ratings of perceived qualification and invitation propensity (H1).

As a specification of H1, one may additionally assume that gender-based preferential selection policies (i.e., preferring female over equally qualified male applicants in areas where women are underrepresented) have a stronger suppressive influence on the invitation propensity than on evaluations of perceived qualification. We will examine this as well.

Moreover, the degree of suppression may depend on evaluators’ contexts. Suppression may be *stronger* in disciplines with more severe women’s underrepresentation because of stronger external normative pressure to appoint female professors in these contexts, which may induce higher compliance with universities’ affirmative action policies and internalization of the gender-based preferential selection norm



(Koch et al., 2015; Williams and Ceci, 2015). For example, analyses based on our data (see Section 4 below) show that 89 % of professors in mathematics/physics and 85 % in the social sciences (economics, sociology, political science) agreed “(very) strongly” with the statement “women are underrepresented in higher academic positions,” but only 71 % in German studies.<sup>3</sup> We therefore expect:

- the female advantage in the ratings of invitation propensity is larger in mathematics/physics than in economics/sociology/political science and especially in German studies (H2a).

A competing mechanism suggests that evaluators may experience different definitions of “fit” based on applicants’ gender, and thus of acceptable prejudice expressions, depending on the gender distribution in their work environment (Miller et al., 2015; Smyth and Nosek, 2015). Eagly and Karau’s (2002) role congruity theory provides a rationale for this mechanism: it assumes that the greater the incongruence between the applicant’s gender and the gender stereotype of a position, the stronger the prejudice. Consistent with this, the meta-analysis by Koch et al. (2015) found that men receive slightly higher ratings than equally qualified women in male-dominated jobs, but they also found no gender differences in ratings for female-dominated and gender-integrated jobs. This mechanism of role (in)congruity may apply to the gender stereotype of disciplines and the level of acceptance of gender-science stereotypes. Thus, suppression of prejudice expression through gender policies may be *weaker* in mathematics/physics than in the other two disciplinary groups. We therefore hypothesize:

- the female advantage in the ratings of both perceived qualification and invitation propensity is smaller in mathematics/physics than in economics/sociology/political science and German studies (H2b).

Both Correll and Ridgeway’s (2003) expectation states theory and Eagly and Karau’s (2002) role congruity theory suggest that individualized information about performance and skills may also help reduce the expression of gender prejudices. However, both theories assume that this is only the case when performance indicators are unambiguous, that is, when they are clearly diagnostic of performance (Koch et al., 2015, p. 139). For example, in academia, collaborative outputs (like co-authorships) may give evaluators more leeway to attribute credit to this work by following gender stereotypes. In line with this, studies for economics have shown that female co-authors receive lower evaluations than male co-authors (Görxhani et al., 2023; Sarsons, 2017; Sarsons et al., 2021). Yet, whether and to what extent co-authorship is associated with gender bias may depend on disciplines’ publication practices. For example, analyses using our data (see Section 4 below) show large disciplinary differences in the incidence of co-authorship: it is “very common” in mathematics/physics (64 %), but much less so in economics/sociology/political science (47 %), and especially in German studies (9 %).<sup>4</sup> In contrast, 43 % of professors in German studies reported that it is “(very) uncommon” in their discipline, compared to only about 5 % in the other two disciplinary groups. All these differences are statistically significant ( $p < 0.05$ ). Against this background, we expect:

- a female co-authorship penalty in the ratings of both perceived qualification and invitation propensity in all three disciplinary groups (H3a), and that this penalty is largest in German studies (H3b).

These two hypotheses are also consistent with the ideas of the

justification-suppression model of prejudice expression. Research has shown that the more respondents suppress the expression of their between-category prejudice (such as gender) because of social norms, the more susceptible they are to less controllable, subtle within-category variation (here, co-authorship judgements) (Marx and Ko, 2019, p. 7).

The extent and overall importance of third-party funding also varies across disciplines (see Petzold and Netz, 2023, for full professor appointments). However, unlike co-authorship, applicants’ third-party funding can be considered to be a fairly *unambiguous* performance indicator. To support the above interpretation of gender bias arising from ambiguity, we also included third-party funding in our study. We do not expect gender differences in its impact on respondents’ evaluation of applicants.

Support for our hypotheses can serve as an indication that gender policies and equality norms at German HEIs suppress expressions of gender-science prejudice—and that the degree of suppression depends on the context (operationalized by the disciplines included in our study).

## 4. Research design, data, and methods

### 4.1. Sample

To test our hypotheses, we conducted a factorial survey experiment in June/July 2020 with professors from the departments of mathematics, physics, economics, political science, sociology and social sciences (combining political science and sociology), and German studies at German universities.<sup>5</sup> The response rate was 22 %. To simulate the situation of incumbent faculty hiring prospective faculty (see Ceci, 2018) and of experienced “real-life” evaluators (Koch et al., 2015, p. 137), we included assistant, associate, and full professors as respondents in our study. They are all eligible to serve on appointment committees in

**Table 1**  
Samples and response rates.

	Gross sample <sup>a</sup>	Response rate (%) <sup>b</sup>	Realized sample	Analytical sample
<b>Total sample</b>				
Total (n)	9454	21.6	2041	1857
Male professors (n)	7546	19.7	1483	1348
Female professors (n)	1906	29.3	558	509
% female professors	20.2		27.3	27.4
Number of vignettes				11,073
<b>Mathematics/physics</b>				
Total (n)	4206	18.6	781	700
Male professors (n)	3687	17.4	643	576
Female professors (n)	517	26.5	137	124
% female professors	12.3		17.5	17.7
Number of vignettes				4186
<b>Economics/sociology/political science</b>				
Total (n)	4117	24.0	990	908
Male professors (n)	3179	22.3	710	649
Female professors (n)	938	29.9	280	259
% female professors	22.8		28.3	28.5
Number of vignettes				5416
<b>German studies</b>				
Total (n)	1131	23.9	270	249
Male professors (n)	680	19.3	131	123
Female professors (n)	451	30.8	139	126
% female professors	39.9		51.5	50.6
Number of vignettes				1471

<sup>a</sup> Collected from university webpages between fall 2019 and spring 2020.

<sup>b</sup> Surveyed in June/July 2020.

<sup>3</sup> The difference between German studies and the other two disciplinary groups is statistically significant ( $p < 0.05$ ).

<sup>4</sup> Including also the “common” category, the differences are smaller but still very substantial: 90 %, 81 %, and 25 %, respectively.

<sup>5</sup> All relevant data are available from GESIS at DOI: <https://doi.org/10.7802/2742>.

**Table 2**  
Profile characteristics of the hypothetical applicants.

Dimensions	Levels	Variable labels (in tables)
<b>Randomly assigned characteristics in applicants' profiles</b>		
<i>Between-subject design</i>		
Applicant's gender	Female [first + last name]	Female applicant
	Male [first + last name]	Male applicant
<i>Within-subject design</i>		
Type of position	Assistant professorship without tenure track	Without tenure track
	Assistant professorship with tenure track	With tenure track
Type of publication record	70 % single- and 30 % co-authored publications	Mostly single-authored
	30 % single- and 70 % co-authored publications	Mostly co-authored
	100 % co-authored publications	All-co-authored
Research collaborations	With renowned scholars	Renowned collaborations
	With scholars of same career level (i.e., postdoctoral researchers)	Early-career collaborations
Research grant (third-party funding)	—	Without a grant
(Reported) family responsibilities	Successful acquisition of a research grant	With a grant
	—	Without parental leave
	6 months parental leave	With parental leave
<b>Fixed characteristics in applicants' profiles</b>		
- Grade of PhD "magna cum laude"		
- 4 years since PhD completion		
- Teaching experience at undergraduate and graduate level		
- Respondent-specific number of peer-reviewed publications required to have a chance of being appointed as an assistant professor		

Germany. About 77 % of the respondents had served at least once on an appointment committee for assistant professors, with no differences between respondents from the different disciplines.

For reasons of statistical power, we used three disciplinary groups in our analyses: mathematics/physics, economics/sociology/political science (including professors from the social science departments), and German studies. Our analytical sample was restricted to respondents with valid information on both dependent variables (see below). We excluded 21 respondents whose year of PhD completion was unknown (we used this date as a control variable for academic experience). These restrictions left us with an *analytical sample* of 1857 respondents (i.e., 91 % of the realized sample).

Table 1 shows that female professors were more likely to participate than male professors, and professors of economics/sociology/political science and German studies were more likely to participate than professors of mathematics/physics. Therefore, we always controlled for respondent gender in our models, and additionally for respondent discipline in the pooled models.<sup>6</sup> Non-random self-selection may reduce the external validity of our study, but due to its experimental design (i.e., randomized assignment of vignettes), it does not affect its internal validity. The issue of external validity is less relevant to our study because our central aim is to test for gender bias and potential explanations (see above).

Although online surveys generally have lower response rates than other survey modes (Daikeler et al., 2020; Manfreda et al., 2008), they are however a more effective and efficient method of collecting data when sample sizes are large compared to other survey modes (Wu et al., 2022). Survey methodology research has also shown that the anonymity of web/online surveys increases true reporting of socially sensitive information and reduces item nonresponse to such questions (Kreuter et al., 2008). Thus, not only our experimental design, but also the web administration of the survey contributes to "true" reporting. In addition, our between-subject design for applicant gender (see below) and the fact that the invitation and the introduction to the survey referred to "recruitment criteria for professorships" (without mentioning "gender") made it difficult to detect our gender-related research question.

Furthermore, given our relatively large analytical sample size and our experimental design, we need only to be reasonably confident that we are not missing certain types of individuals who are pro or con female applicants. We are not aware of missing such specific types of individuals, as our sample includes several disciplines, a wide range of ages (in terms of academic age—from assistant to full professor—and birth cohort) and number of publications of respondents, as well as respondents with/without children.

#### 4.2. Experimental design

Each respondent rated six hypothetical applicants for assistant professorships on a scale of 1 (lowest) to 7 (highest) in two respects relative to their own disciplines: (1) "Would you invite this applicant to an interview?" (invitation propensity) and (2) "Do you think this applicant is qualified for this position?" (perceived qualification). These are the two dependent variables in our analyses.

We provided short descriptions of the hypothetical applicants ("vignettes"). These descriptions varied in terms of the values (called "levels") of the "dimensions" presented. These are our independent variables. The applicant profiles included the applicant's gender (by first name) and information about the type of assistant professorship position, the applicant's type of publication record, research collaborations, third-party funding, and parental leave (see Table 2). Although we do not have hypotheses on all dimensions, variations in these dimensions made the applicant profiles more authentic, thereby increasing the internal validity of our factorial survey.

We also included applicants' number of publications, which is respondent-specific. Before starting to rate the vignettes, the respondents provided their opinion on the approximate number of peer-reviewed publications an applicant should have in their discipline to have a chance of being offered an assistant professorship. This respondent-specific number was then used as a fixed characteristic in the applicant profiles (and included as a control variable in our analyses). Thus, in terms of publication output, our hypothetical applicants should

<sup>6</sup> The statistical power for estimating separate models by respondents' gender within disciplines is too low because of the comparatively small absolute number of female professors, the much smaller sample size for German studies, and the between-subject design for applicant's gender (see Section 4.2).

be eligible for appointment. On average, respondents reported nine required peer-reviewed publications—with mathematics/physics reporting the highest number (13) and economics/sociology/political science the lowest number (7) (controlled for respondents' academic age and gender, the difference narrows to 9 versus 4).

To avoid social desirability, we used a between-subject design for applicant's gender and a within-subject design for all other vignette dimensions (like Carlsson et al., 2021; Ceci, 2018), meaning that respondents rated either male or female vignettes. Walzenbach's (2019) experimental study shows that the between-subject design effectively reduces the risk of social desirability bias (see also Ceci, 2018). Furthermore, the meta-analysis by Koch et al. (2015, p. 138) finds that effect sizes are similar for within-subject and between-subject manipulations of applicant gender. In other words, effect sizes for rating only male or female profiles should not differ from those for competitive ratings of male and female profiles (Koch et al., 2015, p. 131).

Because of the between-subject design for applicant gender, the vignette universe contains 48 possible combinations of all other dimension levels. These 48 vignettes were grouped into eight decks of six vignettes each—identical for male and female applicants. We used the d-optimal blocking approach to generate the decks (Dümler, 2007). The eight decks and applicant gender were randomly assigned to the respondents within each discipline. Respondents were allowed to go back and forth between the six vignettes to avoid censored ratings (Auspurg and Hinz, 2015, p. 214). The order of the vignettes within each deck was varied to control for vignette effects and to ensure the estimation on non-confounded main and interaction effects (Kuhfeld et al., 1994; Su and Steiner, 2020). Our experimental design produced a vignette sample in which dimensions and levels were balanced (i.e., they were rated similarly often) and orthogonal (i.e., they were not correlated with each other), as shown in Tables S1 and S2 in the Online Supplement. This allows for causal interpretations of dimension and level effects on the professors' ratings (Auspurg and Hinz, 2015).

4.3. Analytical strategy

Our two dependent variables are the ratings of (a) invitation propensity and (b) perceived qualification for an assistant professorship on a scale ranging from 1 (lowest) to 7 (highest). We estimated linear models treating the two dependent variables as continuous (xtreg in Stata). Because the data have a multilevel structure, with each professor rating up to six vignettes, we estimated multilevel models with random intercepts (Snijders and Bosker, 2012). The random intercept model chosen allows for unbiased estimates for grouped data by including an individual-specific term for each respondent. The estimation equation is:

$$Y_{ij} = \beta_0 + \beta X_{ij} + \gamma Z_i + \varepsilon_{ij} + \nu_i$$

where

- $Y_{ij}$  = rating on vignette  $j$  of respondent  $i$ ,
- $\beta_0$  = intercept representing the mean average rating of all respondents,
- $X_{ij}$  = vector with the value of the vignette characteristics for vignette  $j$  and respondent  $i$ ,
- $Z_i$  = vector with the characteristics of respondent  $i$  and design variables,
- $\varepsilon_{ij}$  = random error term at the vignette level,
- $\nu_i$  = respondent-specific error term (i.e., a random effect estimated for all observations belonging to a person  $i$  to account for unobserved differences and heterogeneity in the respondents' rating behavior),
- with  $i = 1, \dots, k$  respondents and  $j = 1, \dots, m$  vignettes (Hox et al., 1991).

**Table 3**  
Descriptive statistics of the ratings on invitation propensity and perceived qualification.

	Mean ratings (rating points)	(SD)	Median ratings (rating points)	Rating point shares (%)		
				1–3	4–5	6–7
<b>Mathematics/physics</b>						
Invitation propensity						
Male applicants	5.5	(1.3)	6	7.3	35.7	57.0
Female applicants	5.7	(1.5)	6	8.4	26.3	65.3
Perceived qualification						
Male applicants	5.4	(1.3)	6	7.9	40.1	52.1
Female applicants	5.4	(1.4)	6	9.0	34.6	56.4
<b>Economics/sociology/political science</b>						
Invitation propensity						
Male applicants	5.2	(1.5)	5	13.3	37.7	49.0
Female applicants	5.4	(1.4)	6	10.5	32.1	57.4
Perceived qualification						
Male applicants	5.1	(1.5)	5	13.6	40.4	46.0
Female applicants	5.3	(1.4)	6	10.9	35.6	53.5
<b>German studies</b>						
Invitation propensity						
Male applicants	4.6	(1.7)	5	23.7	42.3	33.9
Female applicants	4.9	(1.7)	5	20.6	35.4	43.9
Perceived qualification						
Male applicants	4.5	(1.6)	5	26.0	44.0	30.0
Female applicants	4.8	(1.6)	5	21.0	37.7	41.3

SD = standard deviation. Rating scale from 1 = low to 7 = high.

The predicted values in our analyses on the basis of these multilevel linear models are not below 1 and not above 7.<sup>7</sup> We also estimated multilevel ordered logit models with random intercepts, given the ordinal rating scale (see Online Supplement, Tables S5–S9). We present the multilevel linear models in the article because they are easier to interpret and save space in terms of the table size they require. Overall, our main findings remain the same.

To test hypothesis H1, we estimated the effect of applicants' gender on perceived qualification and invitation propensity in separate regressions for the three disciplinary groups. To test hypotheses H2a and H2b, we used pooled regressions including a cross-level interaction term of applicants' gender with respondents' disciplinary group. For hypotheses H3a and H3b, we estimated separate disciplinary regressions and included the interaction term of applicants' gender and the type of publication record (H3a) and a pooled regression with three-way interaction terms of applicants' gender with applicants' type of publication record and respondents' disciplinary group (H3b).

All regressions include all vignette dimensions and control for respondents' characteristics (i.e., gender, academic age, and also discipline in pooled models), design variables (i.e., the position of vignettes in the decks and a dummy for all vignettes having the same values), and the respondent-specific number of required publications (see Table 2).

5. Results

Table 3 presents descriptive findings. The mean ratings, ranging from 4.5 to 5.7, and the median ratings of 5 to 6, indicate left-skewed distributions toward higher ratings for male and female applicants for both perceived qualification and invitation propensity (see also Appendix, Fig. A1). This skewness is due to the fact that all hypothetical applicants were formally eligible in terms of the required number of

<sup>7</sup> Residual correlations in these models are very small, ranging from 0.110 to 0.128.



**Table 4**

Multilevel linear regressions: Invitation propensity and perceived qualification by respondents' disciplinary group (in rating points).

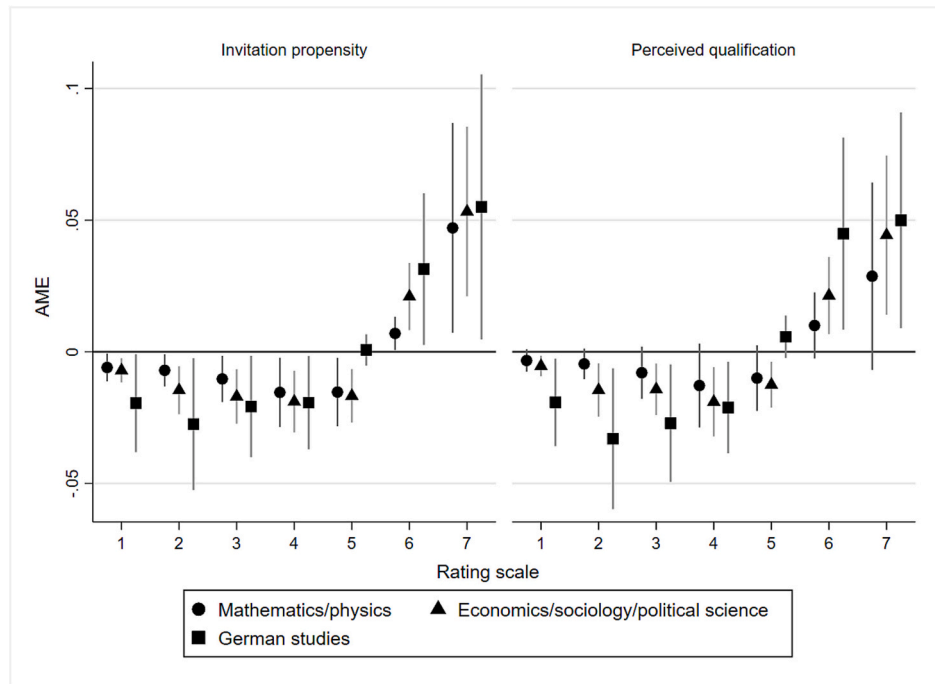
Variables	Mathematics/physics		Economics/sociology/political science		German studies	
	Invitation	Qualified	Invitation	Qualified	Invitation	Qualified
Applicant's gender (ref. male): female	0.117 (0.084)	0.080 (0.080)	0.240** (0.079)	0.223** (0.078)	0.382* (0.173)	0.406* (0.169)
Type of assistant professor position (ref. w/t): with tenure track	−0.011 (0.016)	−0.025 (0.017)	−0.023 (0.016)	−0.023 (0.016)	−0.032 (0.034)	−0.045 (0.033)
Type of publication record (ref. mostly co-authored)						
Mostly single-authored	0.229** (0.026)	0.255** (0.027)	0.351** (0.023)	0.336** (0.023)	0.500** (0.051)	0.485** (0.051)
All co-authored	−0.524** (0.030)	−0.560** (0.031)	−0.443** (0.027)	−0.453** (0.027)	−0.529** (0.054)	−0.560** (0.055)
Research grant (ref. w/t): with a grant	0.377** (0.023)	0.426** (0.024)	0.321** (0.020)	0.327** (0.020)	0.448** (0.045)	0.459** (0.045)
Research collaborations (ref. early-career): Renowned collaborations	0.111** (0.019)	0.116** (0.020)	0.092** (0.016)	0.092** (0.017)	0.198** (0.039)	0.181** (0.037)
Parental leave (ref. w/t): with parental leave	0.069** (0.020)	0.085** (0.021)	0.115** (0.017)	0.110** (0.017)	0.091* (0.042)	0.138** (0.041)
<i>Control variables</i>						
Respondent's gender (ref. male): female	0.224* (0.110)	0.211* (0.103)	0.096 (0.091)	0.126 (0.087)	0.203 (0.192)	0.125 (0.183)
Respondent's academic age (years since PhD)	−0.018 (0.013)	−0.007 (0.012)	−0.012 (0.013)	−0.014 (0.012)	−0.037 (0.027)	−0.031 (0.025)
Respondent's academic age squared	0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Position of vignette within deck	−0.016* (0.007)	−0.022** (0.008)	−0.007 (0.007)	−0.004 (0.007)	−0.031* (0.015)	−0.034* (0.015)
Respondent-specific number of peer-review publications	0.000 (0.000)	0.001* (0.001)	−0.007 <sup>+</sup> (0.004)	−0.007 (0.004)	−0.011* (0.007)	−0.007 (0.006)
Missing information on resp.-specific number	−0.937** (0.239)	−0.672** (0.218)	−0.297** (0.110)	−0.383** (0.107)	−0.623** (0.173)	−0.583** (0.170)
Equal ratings of all vignettes (no = 0, yes = 1)	0.866** (0.115)	0.530** (0.135)	0.556** (0.132)	0.472** (0.139)	0.524 (0.387)	0.489 (0.408)
Intercept	5.548** (0.184)	5.361** (0.177)	5.180** (0.160)	5.200** (0.158)	5.171** (0.418)	4.992** (0.391)
n vignettes	4186	4174	5416	5403	1471	1464
n respondents	700	700	908	908	249	249
Wald $\chi^2$	707.4	749.4	827.1	828.5	352.3	367.5
Prob > $\chi^2$	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
R <sup>2</sup> within	0.247	0.269	0.237	0.237	0.293	0.312
R <sup>2</sup> between	0.135	0.077	0.057	0.068	0.103	0.088
R <sup>2</sup> overall	0.168	0.142	0.107	0.112	0.161	0.161

Robust standard errors in parentheses, \*\*  $p < 0.01$ , \*  $p < 0.05$ , <sup>+</sup>  $p < 0.1$  (two-sided tests). Missing respondent-specific number of required peer-review publications (by 130 respondents) and reported numbers below 3 were replaced by 3 in all respondent's vignettes (236 respondents).

publications (see Section 4.2).

A female advantage or at least equal mean ratings can be observed in all disciplinary groups: In mathematics/physics, the mean rating for

invitation propensity is 0.2 points higher for female applicants (this corresponds to 13 % of one standard deviation, hereafter SD), and the same for male and female applicants for perceived qualification; female



**Fig. 2.** Gender differences in invitation propensity and perceived qualification based on multilevel ordered logit regressions (AME, percentage points). Interpretation example “invitation propensity in mathematics/physics”: Female applicants are 4.7 percentage points more likely to be rated 7 and about 1.5 percentage points less likely to be rated 5 than male applicants. For full models, see Online Supplement, Table S5. All regressions include the other vignette dimensions and all control variables (see Table 4).

**Table 5**

Multilevel linear regressions: Invitation propensity and perceived qualification, including interaction term with applicants' gender and disciplinary group (in rating points).

Variables	Invitation propensity		Perceived qualification	
Applicant's gender (ref: male): female	0.215** (0.055)	0.124 (0.083)	0.192** (0.053)	0.077 (0.079)
Disciplinary group (ref: Mathematics/physics)				
Economics/sociology/political science	−0.257** (0.060)	−0.322** (0.080)	−0.181** (0.058)	−0.259** (0.080)
German studies	−0.641** (0.102)	−0.751** (0.135)	−0.589** (0.095)	−0.742** (0.128)
<i>Interaction terms</i>				
Female applicant * Economics/sociology/political science		0.125 (0.116)		0.150 (0.112)
Female applicant * German studies		0.219 (0.190)		0.306 <sup>+</sup> (0.181)
Intercept	5.473** (0.119)	5.520** (0.121)	5.359** (0.114)	5.420** (0.117)
n vignettes	11,073		11,041	
n respondents	1857		1857	
Wald chi <sup>2</sup>	1900	1905	1928	1935
Prob > chi <sup>2</sup>	<0.001	<0.001	<0.001	<0.001
R <sup>2</sup> within	0.245	0.245	0.257	0.257
R <sup>2</sup> between	0.116	0.117	0.101	0.103
R <sup>2</sup> overall	0.152	0.152	0.145	0.146

Robust standard errors in parentheses, \*\*  $p < 0.01$ , \*  $p < 0.05$ , <sup>+</sup>  $p < 0.1$  (two-sided tests). All models additionally include the other vignette dimensions and all control variables (see Table 4).

**Table 6**

Multilevel linear regressions: Invitation propensity, including ratings for perceived qualification (in rating points).

Variables	Mathematics/physics	Economics/sociology/ political science	German studies
Applicant's gender (ref: male): female	0.108* (0.043)	0.042 (0.038)	−0.015 (0.062)
Respondent's rating of perceived qualification (rating scale 1–7)	0.797** (0.016)	0.849** (0.013)	0.913** (0.013)
Type of assistant prof. position (w/t): with tenure track	0.009 (0.012)	−0.001 (0.010)	−0.001 (0.021)
Type of publication record (ref. mostly co-authored) Mostly single-authored	0.024 (0.015)	0.065** (0.013)	0.051 <sup>+</sup> (0.026)
All co-authored	−0.078** (0.017)	−0.058** (0.016)	−0.020 (0.026)
Research grant (ref: w/t): with a grant	0.036* (0.015)	0.042** (0.013)	0.042* (0.021)
Research collaborations (ref. early-career): Renowned collaborations	0.019 (0.013)	0.011 (0.010)	0.023 (0.020)
Parental leave (ref. w/t): with parental leave	0.002 (0.012)	0.023* (0.010)	−0.029 (0.021)
Intercept	1.274** (0.136)	0.760** (0.102)	0.636** (0.198)
n vignettes	4,171	5,389	1,457
n respondents	700	908	249
Wald chi <sup>2</sup>	4,273	6,429	7,152
Prob > chi <sup>2</sup>	<0.001	<0.001	<0.001
R <sup>2</sup> within	0.734	0.760	0.837
R <sup>2</sup> between	0.748	0.786	0.890
R <sup>2</sup> overall	0.741	0.781	0.872

Robust standard errors in parentheses, \*\*  $p < 0.01$ , \*  $p < 0.05$ , <sup>+</sup>  $p < 0.1$  (two-sided tests). All models additionally include all control variables (see Table 4).

applicants received on average higher ratings for both invitation propensity and perceived qualification in economics/sociology/political science (0.2 points or 14 % of one SD) and in German studies (0.3 points or 19 % of one SD). With very few exceptions, female applicants receive higher mean ratings for almost all vignettes (see Appendix, Fig. A2). These gender differences in mean ratings translate into women being less likely to receive low(er) ratings and more likely to receive high ratings (6–7) than men (see Table 3).

Using two-sample *t*-tests, the gender differences in the mean scores presented in Table 3 are significant ( $p < 0.05$ ) in all three disciplinary groups for invitation propensity, and significant in economics/

sociology/political science and German studies (in mathematics/physics only with  $p < 0.1$ ) for perceived qualification. Two-sample Wilcoxon rank-sum (Mann-Whitney) tests for ordinal data confirmed the statistical significance ( $p < 0.05$ ) of the gender differences for both outcomes in all three disciplinary groups.

However, the figures presented in Table 3 are not yet comparable across disciplines because the composition of disciplines differs by respondents' gender and academic age. We therefore test hypothesis H1 by using regressions, presented in Table 4. In support of H1, we find a female advantage for both invitation propensity and perceived qualification for applicants in economics/sociology/political science (of 0.240

**Table 7**  
Gender differences in invitation propensity by perceived qualification (AME in rating points).

	Mathematics/physics	Economics/sociology/ political science	German studies
Gender difference (ref: male applicants) among those with			
Low perceived qualification (ratings 1–3)	0.085 (0.181)	−0.185 (0.130)	0.127 (0.155)
Medium perceived qualification (ratings 4–5)	0.133 <sup>+</sup> (0.068)	0.129** (0.058)	0.109 (0.103)
High perceived qualification (ratings 6–7)	0.129* (0.051)	0.118** (0.045)	0.008 (0.089)

Robust standard errors in parentheses, \*\*  $p < 0.01$ , \*  $p < 0.05$ , <sup>+</sup>  $p < 0.1$  (two-sided tests). Based on multilevel linear regressions with random intercept: likelihood of being invited, including respondents' rating scores for perceived qualification (see Online Supplement, Table S3). All models additionally include the other vignette dimensions and all control variables (see Table 4).

and 0.223 rating points, respectively) and German studies (of 0.382 and 0.406 rating points, respectively).<sup>8</sup> For mathematics/physics, the effect of applicant's gender also indicates a female advantage, albeit smaller and statistically insignificant. Fig. 2, based on multilevel ordered logit models, shows the gender differences in the ratings ultimately received by male and female applicants. We see that female applicants were more likely to receive high rating scores (6 and 7) and less likely to receive low (er) rating scores than male applicants. The differences between invitation propensity and perceived qualification are small and not significant.

Table 4 further shows that all other vignette dimensions influence the ratings in the expected way in all three disciplinary groups.<sup>9</sup> The type of publication record is the most influential criterion (as indicated by the differences in the coefficients for mostly single- vs. all co-authored publications). The second strongest predictor is having received a research grant. In addition, applicants with research collaborations with renowned scholars receive a slightly higher rating than those with collaborations with scholars at the same level (postdocs),<sup>10</sup> as do applicants with parental leave than those without parental leave. Note, however, that those with and without parental leave were assigned the same number of publications (i.e., the same respondent-specific number of required publications as a fixed vignette profile characteristic). Interestingly, the type of assistant professorship (tenure-track/non-tenure-track) does not matter for the evaluations (although it is of course important for applicants in real life and may influence their application behavior). The interaction between type of assistant professorship and applicants' gender is also negligible and insignificant (results not shown).

The main research question of our study, however, is whether the magnitude of the female advantage differs across disciplinary groups. Hypothesis H2a predicted a larger female advantage in mathematics/physics because of stronger prejudice suppression due to higher external

normative pressure. In contrast, H2b expected the female advantage to be smaller in mathematics/physics because of a weaker suppression due to role incongruity. To test these two hypotheses, Table 5 reports the results of pooled regressions with interaction terms between applicants' gender and disciplinary groups. The interaction terms are not significant, though of notable size and indicating a smaller female advantage in mathematics/physics. The interaction effect for perceived qualification ratings comes close to the conventional significance level (with  $p < 0.1$ ) for the difference between mathematics/physics and German studies. Thus, the direction of the interaction terms tentatively supports hypothesis H2b.

Table 5 assumes an equal qualification of male and female applicants on the basis of the vignette dimensions when estimating the invitation propensity for a job interview. However, we also found a female advantage in the ratings of perceived qualifications; that is, male and female applicants' perceived qualifications differ despite identical profiles (at least in economics/sociology/political science and German studies). While the primary aim of gender policies is to prevent biased perceptions of competence and to prefer women when they are equally qualified, they may also indirectly increase the “perceived qualification” of female applicants. Ostensibly neutral criteria for defining “merit” and the “ideal candidate” often favor male-typical career trajectories and achievements, resulting in institutional discrimination (Orupabo and Mangset, 2022). This occurs when such criteria fail to account for gender differences in life courses, teaching and administrative workloads, and recognition of research contributions—topics that are addressed in anti-bias training and appointment guidelines. Consequently, awareness of these disparities may lead evaluators to rate women's “equal” performance more favorably in terms of perceived qualifications. Moreover, respondents may rationalize their higher invitation ratings by attributing them to the applicant's “merits.” Tables 6 and 7 therefore include the ratings of applicants' perceived qualifications in the invitation regression—as a subjective measure of “equal qualification.”

The strong impact of perceived qualification on invitation propensity in Table 6 suggests that respondents make merit-based judgements, at least based on their perception of qualifications (which are themselves merit-based, as is evident in Table 4 above). Table 6 further shows that in economics/sociology/political science and German studies, the female advantage in invitation propensity, net of perceived qualification, is considerably smaller than in Table 4. This difference suggests that the female advantage in invitation propensity observed in Table 4 is indeed partly due to a higher perceived qualification of female than of male applicants with the same profile. Moreover, the female advantage in invitation propensity is now larger (and statistically significant only) in mathematics/physics than in the other two disciplinary groups. In other words, only in mathematics/physics do female applicants receive slightly higher invitation ratings over and above their perceived qualification. This finding supports hypothesis H2a (in contrast to the findings of Table 5 above) and may indicate that the norm of gender-based preferential selection plays a greater role here than in the other two disciplinary groups.

While Table 6 reports gender differences averaged across all perceived qualification ratings, Table 7 does so for three (grouped)

<sup>8</sup> As an illustration, the average of 0.240 rating points could mean, for example, that 24 % of the female vignettes received a one-point higher rating than male vignettes with the same profile. Of course, if x% of women received a one-point lower rating than men, this would mean that x% + 24 % of women received a one-point higher rating than men to keep the average at 0.24. It is also possible (but less likely) that some women received a 2-point higher rating, which would reduce the number (%) of women with higher ratings, and so on.

<sup>9</sup> Pooled regressions with two-way interactions between these other vignette dimensions and disciplinary group indicate a significantly larger effect of mostly single-authored publications in economics/sociology/political science and German studies than mathematics/physics, and smaller effects of all publications being co-authored in economics/sociology/political science and of research grants in economics/sociology/political science than in the other two disciplinary groups. All other differences between the disciplinary groups are insignificant. (Results are not shown.)

<sup>10</sup> However, as shown in the study by Jungbauer-Gans and Gross (2013), the impact of “renowned scholars” may depend on whether they are the mentors or not. The study found that publishing with the mentor tends to be a disadvantage in mathematics, but an advantage in law, and has no effect in sociology.

**Table 8**  
Multilevel linear regressions: Invitation propensity and perceived qualification, including interaction terms, by respondents' disciplinary group (in rating points).

Variables	Invitation propensity						Perceived qualification					
	Mathematics/physics		Economics/sociology/political science		German studies		Mathematics/physics		Economics/sociology/ political science		German studies	
Applicant's gender (ref: male): female	0.078 (0.089)	0.170 <sup>+</sup> (0.087)	0.226** (0.086)	0.229** (0.080)	0.433* (0.187)	0.367* (0.173)	0.065 (0.085)	0.095 (0.083)	0.218** (0.084)	0.208** (0.079)	0.523** (0.183)	0.394* (0.168)
Type of publication record (ref. mostly co-authored)												
Mostly single-authored	0.209** (0.040)	0.229** (0.026)	0.349** (0.033)	0.351** (0.023)	0.556** (0.076)	0.500** (0.051)	0.240** (0.042)	0.255** (0.027)	0.342** (0.033)	0.336** (0.023)	0.568** (0.077)	0.485** (0.051)
All co-authored	−0.565** (0.043)	−0.524** (0.030)	−0.462** (0.038)	−0.443** (0.027)	−0.510** (0.078)	−0.528** (0.054)	−0.568** (0.045)	−0.560** (0.031)	−0.466** (0.036)	−0.453** (0.027)	−0.474** (0.073)	−0.560** (0.055)
Research grant (ref: w/t): with a grant	0.377** (0.023)	0.431** (0.035)	0.321** (0.020)	0.310** (0.027)	0.448** (0.045)	0.433** (0.067)	0.427** (0.024)	0.442** (0.036)	0.327** (0.020)	0.312** (0.027)	0.459** (0.045)	0.447** (0.067)
<i>Interaction terms</i>												
Applicant's gender * mostly single-authored	0.039 (0.053)		0.003 (0.046)		−0.114 (0.103)		0.030 (0.055)		−0.011 (0.045)		−0.174 <sup>+</sup> (0.102)	
Applicant's gender * all co-authored	0.078 (0.060)		0.039 (0.055)		−0.039 (0.108)		0.015 (0.062)		0.025 (0.054)		−0.179 <sup>+</sup> (0.109)	
Applicant's gender * with research grant		−0.105* (0.046)		0.022 (0.040)		0.031 (0.090)		−0.030 (0.048)		0.029 (0.039)		0.025 (0.090)
Intercept	5.569** (0.186)	5.523** (0.185)	5.187** (0.161)	5.185** (0.160)	5.150** (0.420)	5.178** (0.421)	5.368** (0.179)	5.353** (0.177)	5.202** (0.159)	5.207** (0.157)	4.941** (0.394)	4.998** (0.393)
n vignettes	4186		5416		1471		4174		5403		1464	
n respondents	700		908		249		700		908		249	
Wald chi <sup>2</sup>	711	710	834	828	354	354	749	750	831	829	374	371
Prob > chi <sup>2</sup>	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
R <sup>2</sup> within	0.247	0.248	0.237	0.237	0.294	0.294	0.269	0.269	0.237	0.237	0.314	0.312
R <sup>2</sup> between	0.135	0.135	0.057	0.057	0.102	0.102	0.077	0.077	0.0685	0.069	0.088	0.088
R <sup>2</sup> overall	0.168	0.168	0.107	0.107	0.161	0.161	0.142	0.142	0.112	0.112	0.162	0.161

Robust standard errors in parentheses, \*\* p < 0.01, \* p < 0.05, + p < 0.1 (two-sided tests). All models additionally include the other vignette dimensions and all control variables (see Table 4).

levels of perceived qualification. We find quite different patterns for the disciplinary groups. In mathematics/physics, we see a clear and significant female advantage among applicants with high (6–7) qualification ratings, and close to significance among those with medium (4–5) qualification ratings ( $p = 0.051$ ). In economics/sociology/political science, the female advantage is also significant for the medium and high qualification group, but women receive significantly lower ratings than men in the low (1–3) qualification group (see also Online Supplement, Table S3). In German studies, the gender differences are not statistically significant and close to zero for the high (6–7) qualification group. We also calculated average marginal effects of being a female applicant based on our multilevel order logit models (see Appendix, Fig. A3). They support these linear findings. The differences between the disciplinary groups are not significant, however, as indicated by insignificant three-way interaction terms in pooled regressions.

Summarizing the results from Tables 5–7, we find that hypothesis H2b—larger female advantage in German studies and economics/sociology/political science—is tentatively supported when “equal qualification” is defined using the vignette variables (Table 5). In contrast, when “equal qualification” is defined by respondents’ perceived qualification, hypothesis H2a—larger female advantage in mathematics/physics—is supported (Tables 6 and 7). Overall, however, the impact of the observed female advantage by equal perceived qualifications on the invitation propensity is actually very small (see Appendix, Figs. A3 and A4).

These different results for applicants’ qualifications defined by vignette profile vs. “perceived qualification” could be partly due to the fact that vignette dimensions are perceived differently for male and female applicants—as expected by hypotheses H3a and H3b. We hypothesized that the type of information about the applicant’s performance may influence the strength of prejudice suppression in evaluations. Accordingly, we expected that female applicants would be penalized more strongly for co-authorship (as ambiguous information) than male applicants (H3a), and that this gender bias would be largest in German studies (H3b). We therefore estimated separate disciplinary regressions including interaction terms between applicants’ gender and type of publication record.

Table 8 shows that most of these interaction effects are small in size and statistically insignificant. Only the interaction terms of applicants’ gender and type of publication record for perceived qualification in German studies are close to the conventional significance level (with  $p < 0.1$ ). They indicate a smaller premium for mostly single-authored publications among female applicants ( $0.568 + (-0.174) = 0.394$ ) than among male applicants (0.568) and a larger penalty for all co-authored publications among female than among male applicants ( $-0.474 + (-0.179) = -0.653$  vs.  $-0.474$ ). This German studies’ particularity is statistically significant (tested with pooled models including three-way interaction terms, results not shown). Thus, hypothesis H3a is not supported, while H3b is supported but also extended to a gender-biased premium for mostly single-authored publications. Since there is no doubt about who the author is for mostly single-author publications, the lower premium for female applicants could be due to ambiguity about the content of the publication and/or the quality of the peer-reviewed journals, for which we did not provide information in the vignettes.

We did not expect gender differences in the effect of research grants. Our results in Table 8 largely support this expectation. We see only one statistically significant interaction for the invitation propensity in mathematics/physics. Here, the premium for a research grant seems to be smaller among female applicants ( $0.431 + (-0.105) = 0.326$ ) than among male applicants (0.431). Still, female applicants with a research grant get similar high ratings as equally qualified male applicants ( $5.569 + 0.431 + 0.170 + (-0.105) = 6.065$  vs.  $5.569 + 0.431 = 6.000$ ). This suggests that respondents in mathematics/physics do not apply “double (stricter) standards” when evaluating female applicants.

We did not specify the amount of third-party funding, but only stated

that the “applicant successfully obtained a research grant.” After rating of the vignettes, we asked the respondents the following question: “Some of the applicants you rated have acquired a research grant. What amount of third-party funding did you have in mind when making your evaluations?” Not surprisingly, the amounts vary considerably across the three disciplinary groups (see Online Supplement, Table S4). More interestingly, respondents in mathematics/physics and German studies expected smaller amounts from female than male applicants (about 17,000 and 12,000 euros less, respectively), while the opposite is true for economics/sociology/political science (about 12,000 euros more).<sup>11</sup> This finding supports the idea that more ambiguous information is open to gender bias—but in both directions and, interestingly, not resulting in corresponding gender differences in the ratings (see Table 8).

## 6. Discussion and conclusions

Despite the fact that the transition to “assistant professorships is a key juncture in understanding the problem of women’s underrepresentation” (Williams and Ceci, 2015, p. 5360), research on gender bias in assistant professor appointments is still very limited. Similarly, existing research rarely compares across disciplines. Against this background, our study examined the incidence of, and disciplinary differences in, gender bias in *assistant professor* recruitment using a factorial survey experiment conducted in 2020 with professors at German universities from very different disciplines—mathematics/physics, economics/sociology/political science, and German studies.

Drawing on Crandall and Eshleman’s (2003) justification-suppression model of prejudice expression, we argued that the relatively strong normative climate of gender equality in Germany and affirmative action policies at German universities (see Section 2) may contribute to suppressing the expression of gender-science stereotypes, but that the degree of suppression may vary across disciplinary contexts. Our analyses provide evidence that female applicants for assistant professorships do indeed receive higher ratings than male applicants, both in terms of being perceived as qualified for the position and being invited for an interview—similar to Carey et al. (2020) and Williams and Ceci (2015), but contrary to Steinpreis et al. (1999) and Ooms et al. (2019). This female advantage, however, does not mean that professors in Germany are prejudiced *against men*. Rather, in line with Crandall and Eshleman’s model, we argue that they are less likely to express their prejudice against women. This is not because our factorial survey elicited gendered responses; this was avoided by the between-subject design, as respondents rated either male or female applicants.

Based on the observed differences between the disciplines included in our study, we found stronger support for the justification-suppression model of prejudice expression than for Eagly and Karau’s (2002) role incongruity model. The idea of the latter is that the greater the incongruence between the applicant’s gender and the gender stereotype of a discipline is—in our study exemplified by mathematics and physics—the stronger the gender prejudices. Accordingly, one should expect a male advantage, or at least a smaller female advantage, in mathematics/physics than in economics/sociology/political science or German studies. In contrast, we found a larger female advantage in the invitation propensity in mathematics/physics than in the other two disciplines when “equal qualification” is defined by respondents’ perceptions of applicants’ qualification. This seems to suggest that the normative pressure to comply with gender-based preferential selection is greater in disciplines with extremely low proportions of women among associate/full professors and comparatively low proportions among assistant professors.

Furthermore, we showed that suppression of prejudice expression may depend on the (un)ambiguity of information about applicants’

<sup>11</sup> The gender differences in German Studies and economics/sociology/political science reached  $p < 0.1$  (controlled for respondents’ gender and age).



performance—and also differs by discipline. In mathematics/physics and economics/sociology/political science, we found no significant gender differences concerning single vs. co-authorship. Only in German studies did we observe that female applicants received a smaller premium than male applicants for mostly single-author publications and a larger penalty for all co-authorship in the perceived qualification ratings. However, for mathematics/physics, we observed that having received a research grant (without mentioning the grant amount) generates a smaller premium for female than male applicants (mainly due to a larger female advantage among applicants without a research grant). These findings seem to support the idea that ambiguous information leaves room for gender bias, and that suppression of expression of between-category prejudice due to social norms may increase susceptibility to less controllable, subtle within-category variation (Marx and Ko, 2019, p. 7).

From a policy perspective, our study contributes to the growing literature suggesting that the main cause of women's underrepresentation among professors today is not gender bias in the appointment process, as the prevailing narrative claims, but rather women's lower propensity to apply for professorships, which is also evident at the assistant professor level (see Section 2, Fig. 1). Our finding of a female advantage does not imply that affirmative action policies are not important in the process of overcoming women's underrepresentation. On the contrary, these policies are important in suppressing prejudice expression and contribute to the female advantage we find in this study. However, the female advantage we observed is too small to explain the large increase in the proportion of women from “applications” to “offers” in Fig. 1 (see Section 2). Thus, this increase, as well as the differences between disciplinary groups, is most likely due to a higher quality of female applicants and/or multiple offers to the same women.

The main policy implication of our study is that, in addition to continuing affirmative action practices related to appointment processes, further progress in women's representation among professors will require more effective policies to increase the proportion of female applicants—and this across disciplines. Women's lower application propensity originates both inside and outside academia. Within academia, for example, women may perceive themselves as less qualified than men due to gender differences in experiences of collaboration and co-authorship (Ross et al., 2022) or in teaching evaluations (Ceci et al., 2023). Barriers to women's careers are also rooted in gendered responsibilities for childcare and housework, as well as gendered decisions about geographic mobility within couples (Solga and Rusconi, 2007; Vohlídalová, 2014), which conflict with the supposedly gender-neutral requirements of academic careers.

Finally, our study suggests avenues for further research. Similar to other factorial surveys, the external validity of our results may be affected by the hypothetical character of the (low-stake) decision-making situation. However, the anonymous situation for evaluating applicants has been shown to encourage professors to behave as they would in real decision-making situations (Petzold and Netz, 2022). Respondents' self-selection in the sample could of course be another source limiting the external validity of our results, which we cannot rule out.

Moreover, our sample size of female professor respondents was too small to examine single disciplinary differences by respondent gender—especially given the between-subject design and the small number of female professors in the real world, especially in mathematics and physics. Thus, interesting avenues for future research include exploring disciplinary differences in the strength of prejudice suppression between male and female professors (“evaluators”), extending the study of suppression of prejudice expression in appointment decisions to other countries and additional disciplines, and examining prejudice suppression more directly.

### CRediT authorship contribution statement

**Heike Solga:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Alessandra Rusconi:** Writing – review & editing, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sophie Hofmeister:** Formal analysis.

### 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).

### Funding acknowledgement

This project was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689). Open Access funding was provided by the WZB through the DEAL-Consortium.

### Declaration of competing interest

The authors declare that they have no conflicts of interest.

### Acknowledgements

We thank the editor and two reviewers for the constructive comments. We also thank Klarita Gërxhani, Nevena Kulic, and Fabienne Liehti for their collaboration in designing the study, Oleg Hirsch for his manifold support as student assistant at the WZB, and Monika Hasenbruch, Torben Rauhut, and Melanie Tank from the DZWH for their support in collecting the address data. We thank Adam Reiner and Britta Jurtz for programming the survey instrument and questionnaire. We received valuable feedback from Valentina di Stasio and participants in the WZB workshop series “Experiments in Education and Labor Market Research.”

Appendix A

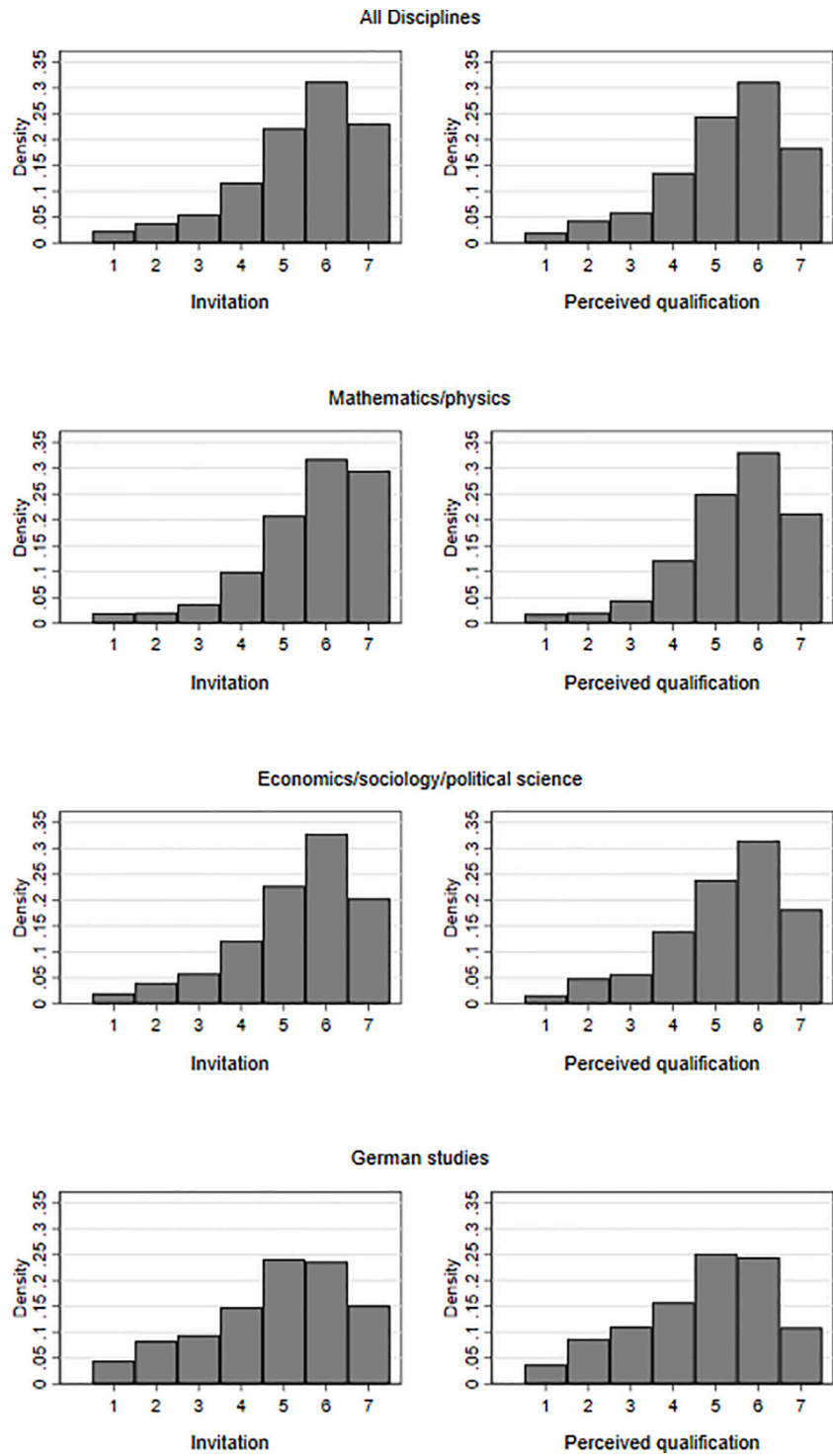


Fig. A1. Distribution of professors' assessments: Rating scores for invitation propensity and perceived qualification.

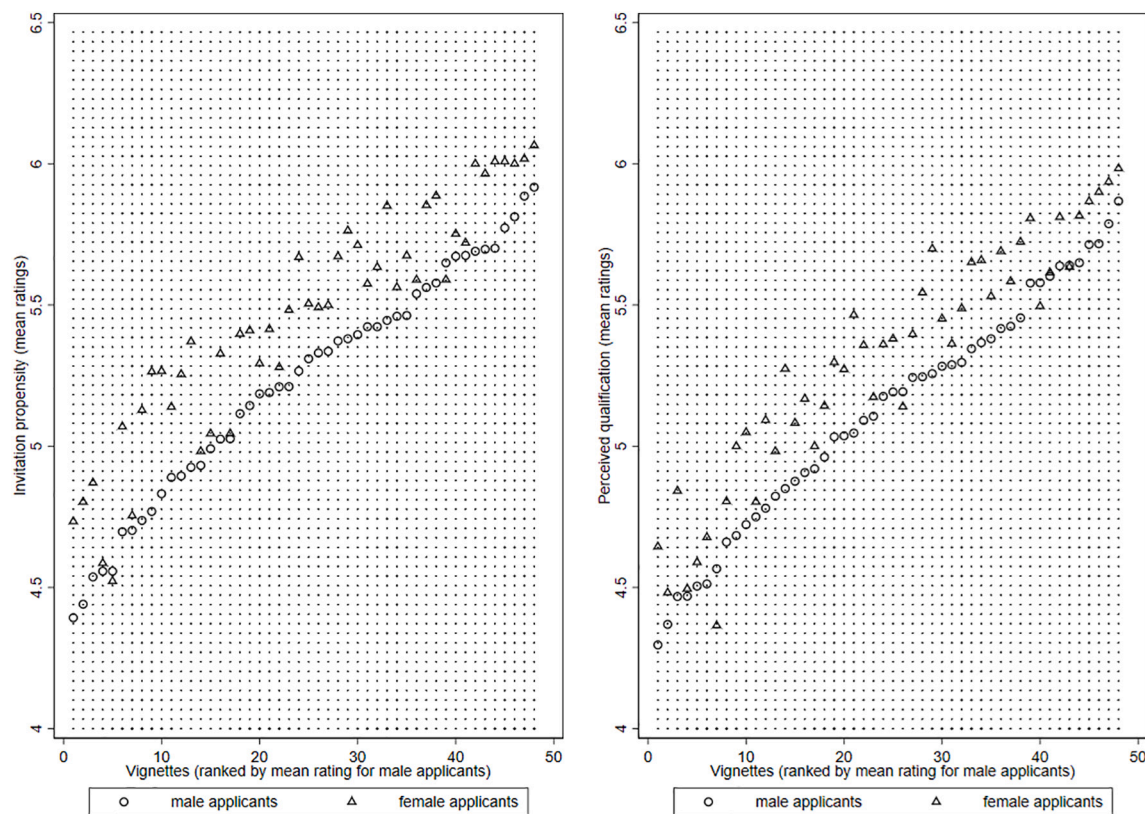


Fig. A2. Mean rating scores of male and female applicants for invitation propensity and perceived qualification, by vignette.

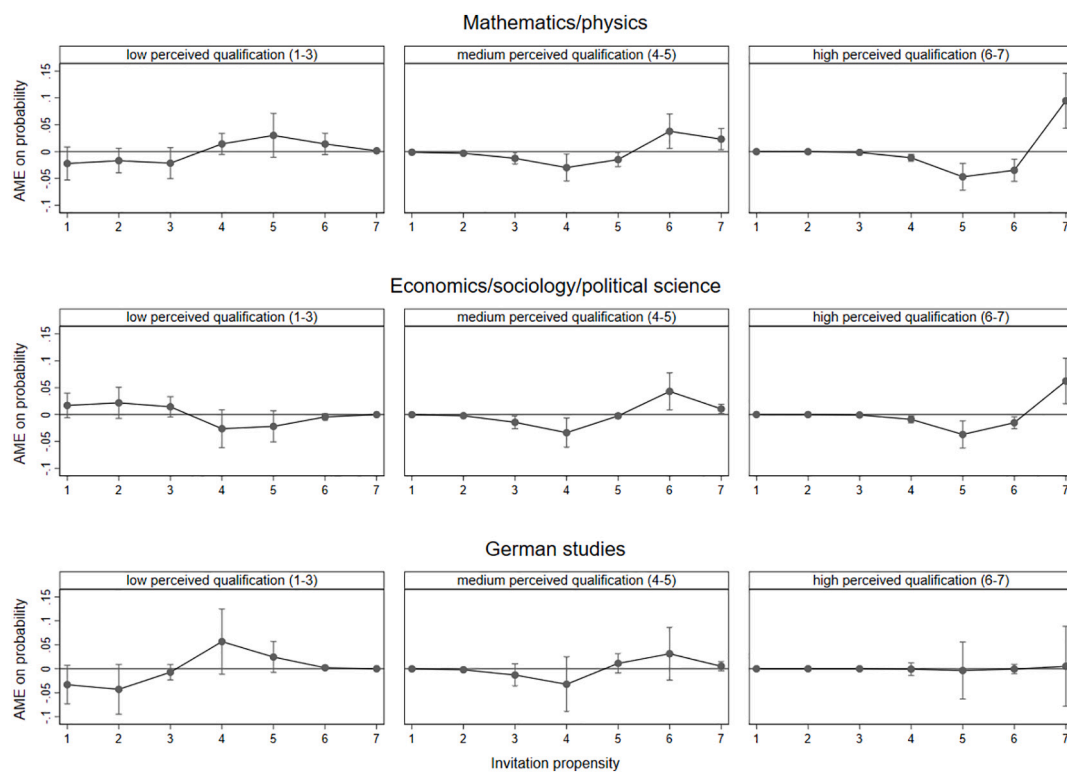
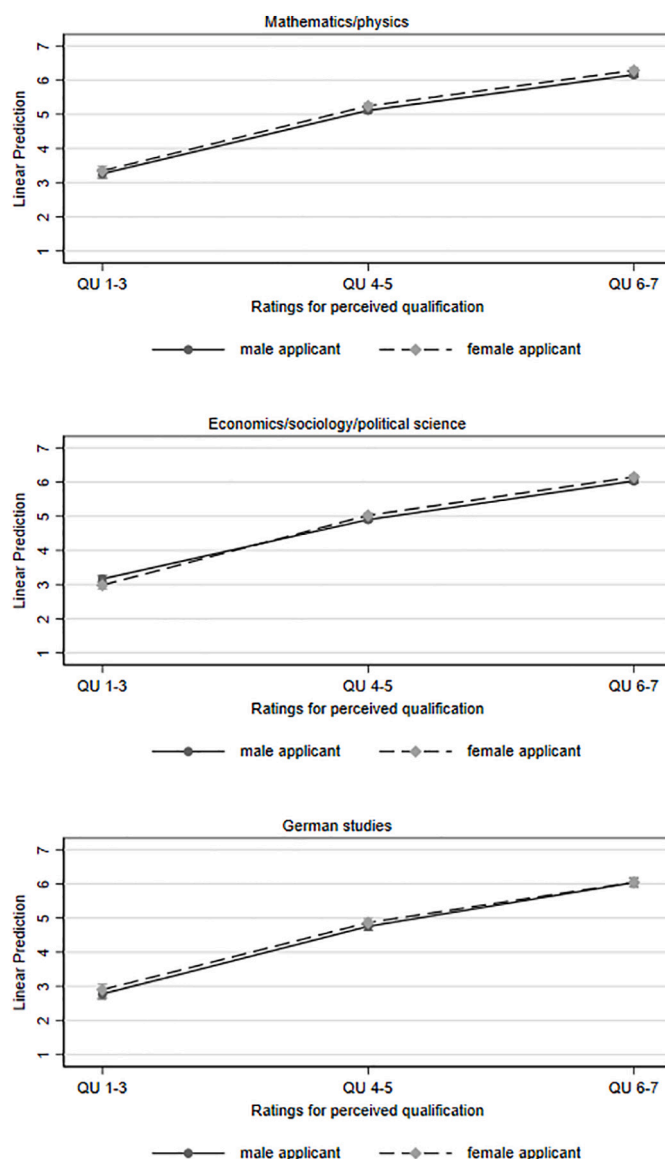


Fig. A3. Average marginal effects (AME) of gender on the probability of invitation propensity ratings by perceived qualification ratings.

Interpretation example “mathematics/physics, high perceived qualification (6–7): Being a female applicant (with perceived qualification 6–7) increases the probability of being rated 7 by 9.5 percentage points and reduces the probability of being rated 5 or 6 by around 5 and 3 percentage points respectively. Based on multilevel ordered logit regressions with random intercept: Invitation propensity, including respondents’ rating scores for perceived qualification (see Online Supplement, Table S8).



**Fig. A4. Predicted ratings of invitation propensity by perceived qualification (in rating points).**

Based on multilevel linear regressions with random intercept: Invitation propensity, including respondents' rating scores for perceived qualification (see Online Supplement, Table S3).

## Appendix B. Online Supplement

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.respol.2024.105170>.

## Data availability

All relevant data are available from GESIS at DOI: <https://doi.org/10.7802/2742>

## References

- Aksnes, D.W., Kahn, S., Reiling, R.B., Ulvestad, M.E.S. 2022. Longitudinal evidence on norwegian PhDs suggests slower progression for women academics but not a leaky pipeline downloaded on 10 April 2022 from [10.31235/osf.io/pvx8q](https://doi.org/10.31235/osf.io/pvx8q).
- Auspurg, K., Hinz, T., 2015. Factorial Survey Experiments. Sage, Thousand Oaks, CA.
- Auspurg, K., Hinz, T., Schneck, A., 2017. Berufungsverfahren als Turniere: Berufungschancen von Wissenschaftlerinnen und Wissenschaftlern. *Z. Soziol.* 46, 283–302.
- Berger, J., Cohen, B.P., Zeldich Jr., M., 1972. Status characteristics and social interaction. *Am. Sociol. Rev.* 37, 241–455.
- Carey, J.M., Carman, K.R., Clayton, K.P., Horiuchi, Y., Htun, M., Ortiz, B., 2020. Who wants to hire a more diverse faculty? Politics, Groups Identities 8, 535–553.
- Carli, L.L., Alawa, L., Lee, Y., Zhao, N., Kim, E., 2016. Stereotypes about gender and science. *Psychol. Women Q.* 40, 244–260.
- Carlsson, M., Finseraas, H., Midtbøen, A.H., Rafnsdóttir, G.L., 2021. Gender Bias in academic recruitment? Evidence from a survey experiment in the Nordic region. *Eur. Sociol. Rev.* 37, 399–310.
- Ceci, S.J., 2018. Women in academic science: experimental findings from hiring studies. *Educ. Psychol.* 53, 22–41.
- Ceci, S.J., Williams, W.M., 2015. Women have substantial advantage in STEM faculty hiring. *Front. Psychol.* 6 article 1532.
- Ceci, S.J., Kahn, S., Williams, W.M., 2021. Stewart-Williams and Halsey argue persuasively that gender bias is just one of many causes of women's underrepresentation in science. *Eur. J. Personal.* 35, 40–44.
- Ceci, S.J., Kahn, S., Williams, M.W., 2023. Exploring gender bias in six key domains of academic science. *Psychol. Sci. Public Interest* 24, 15–73.
- Correll, S.J., Ridgeway, C.L., 2003. Expectation states theory. In: Delamater, J. (Ed.), *Handbook of Social Psychology*. Springer, Boston, pp. 29–51.

- Crandall, C.S., Eshleman, A., 2003. A justification-suppression model of the expression and experience of prejudice. *Psychol. Bull.* 129, 414–446.
- Crosby, F.J., Iyer, A., Sincharoen, S., 2006. Understanding affirmative action. *Annu. Rev. Psychol.* 57, 585–611.
- Daikeler, J., Bošnjak, M., Manfreda, K., 2020. Web versus other survey modes. *Journal of Survey Statistics and Methodology* 8, 513–539.
- Dülmer, H., 2007. Experimental plans in factorial surveys. *Sociol. Methods Res.* 35, 382–409.
- Eagly, A.H., Karau, S.J., 2002. Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598.
- European Commission, 2014. Gender Equality Policies in Public Research. Publications Office of the European Union, Luxembourg.
- European Commission, 2020. A union of equality: gender equality strategy 2020–2025. downloaded 10 April 2022 from. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0152&from=EN>.
- European Commission, 2021. She Figures 2021. Publications Office of the European Union, Brussels.
- Fiske, S.T., Dupree, C., 2014. Gaining trust as well as respect in communicating to motivated audiences about science topics. *PNAS Proceedings of the National Academy of Sciences* 111, 13593–13597.
- Fiske, S.T., Cuddy, A.J.C., Glick, P., Ju, X., 2002. A model of (often mixed) stereotype content. *J. Pers. Soc. Psychol.* 82, 878–902.
- Foschi, M., 2000. Double standards for competence. *Annu. Rev. Sociol.* 26, 21–42.
- Gërkhani, K., Kulic, N., Liechti, F., 2023. Double standards? Co-authorship and gender bias in early-stage academic evaluations. *Eur. Sociol. Rev.* 39, 194–209.
- Gërkhani, K., Kulic, N., Rusconi, A., Solga, H., 2025. Gender bias in evaluating assistant professorship applicants? Evidence from harmonized survey experiments in Germany and Italy. *Soc. Sci. Res.* 126 article 103113.
- GWK, 2019. Chancengleichheit in Wissenschaft und Forschung. In: Materialien der GWK, vol. 69. Gemeinsame Wissenschaftskonferenz, Bonn.
- GWK, 2021. Chancengleichheit in Wissenschaft und Forschung. In: Materialien der GWK, vol. 75. Gemeinsame Wissenschaftskonferenz, Bonn.
- Habicht, I.M., Schröder, M., Lutter, M., 2024. Female advantage in German sociology. Does accounting for the “leaky pipeline” effect in becoming a tenured university professor make a difference? *Soziale Welt*, special issue 26, 407–456. <https://doi.org/10.5771/9783748925590-407>.
- Heilman, M.E., 2012. Gender stereotypes and workplace bias. *Res. Organ. Behav.* 32, 113–135.
- Henningsen, L., Horvath, L.K., Jonas, K., 2022. Affirmative action policies in academic job advertisements. *Sex Roles* 86, 34–48.
- Herschberg, C., Benschop, Y., van den Brink, M., 2018. Precarious postdocs: a comparative study on recruitment and selection of early-career researchers. *Scand. J. Manag.* 34, 303–310.
- Hox, J.J., Kreft, I.G.G., Hermkens, P.L.J., 1991. The analysis of factorial surveys. *Sociol. Methods Res.* 19, 493–510.
- Jungbauer-Gans, M., Gross, C., 2013. Determinants of success in university careers. *Z. Soziol.* 42, 74–92.
- Koch, A.J., D’Mello, S.D., Sackett, P.R., 2015. A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *J. Appl. Psychol.* 100, 128–161.
- Kreuter, F., Presser, S., Tourangeau, R., 2008. Social desirability bias in CATI, IVR, and web surveys. *Public Opin. Q.* 72, 847–865.
- Kuhfeld, W.F., Tobias, R.D., Garratt, M., 1994. Efficient experimental design applications with marketing research applications. *J. Mark. Res.* 31, 545–557.
- LERU, 2018. Implicit Bias in Academia: A Challenge to the Meritocratic Principle and to Women’s Careers. Advice Paper no. 23. League of European Research Universities, Leuven downloaded 12 February 2022 from. <https://www.leru.org/files/Publications/Implicit-bias-in-academia-Full-Paper.pdf>.
- Lutter, M., Habicht, I.M., Schröder, M., 2022. Gender differences in the determinants of becoming a professor in Germany. An event history analysis of academic psychologists from 1980 to 2019. *Res. Policy* 51 article 104506.
- Manfreda, K., Bošnjak, M., Berzelak, J., Haas, I., Vehovar, V., 2008. Web surveys versus other survey modes: a meta-analysis comparing response rates. *Int. J. Mark. Res.* 50, 79–104.
- Marx, D.M., Ko, S.J., 2019. Stereotypes and prejudice. *Oxford Research Encyclopedia of Psychology*. <https://doi.org/10.1093/acrefore/9780190236557.013.307> (online).
- Miller, D.I., Eagly, A.H., Linn, M.C., 2015. Women’s representation in science predicts national gender-science stereotypes. *J. Educ. Psychol.* 107, 631–644.
- National Academies of Sciences, Engineering, and Medicine, 2020. Practices for Addressing the Underrepresentation of Women in Science, Engineering, and Medicine: Opening Doors. The National Academies Press, Washington, DC. <https://doi.org/10.17226/25585>.
- Ooms, W., Werker, C., Hopp, C., 2019. Moving up the ladder. *Stud. High. Educ.* 44, 1268–1289.
- Orupabo, J., Mangset, M., 2022. Promoting diversity but striving for excellence: opening the ‘black box’ of academic hiring. *Sociology* 56, 316–332.
- Petzold, K., Netz, N., 2022. Vignettenexperimente in der Hochschul- und Wissenschaftsforschung. In: Brandt, G., de Vogel, S. (Eds.), *Survey-Methoden in der Hochschulforschung*. VS Springer, Wiesbaden, pp. 153–199.
- Petzold, K., Netz, N., 2023. How do signals of academic performance vary across disciplines? *Soziale Welt*, special issue 26, 490–523.
- Ross, M.B., Glennon, B.M., Murciano-Goroff, R., Berkes, E.G., Weinberg, B.A., Lane, J.I., 2022. Women are credited less in science than men. *Nature* 608, 135–145.
- Sarsons, H., 2017. Recognition for group work: gender differences in academia. *Am. Econ. Rev.* 107, 141–145.
- Sarsons, H., Gërkhani, K., Reuben, E., Schram, A., 2021. Gender differences in recognition for group work. *J. Polit. Econ.* 129, 101–147.
- Schröder, M., Lutter, M., Habicht, I.M., 2021. Publishing, signaling, social capital, and gender: determinants of becoming a tenured professor in German political science. *PLoS One* 16 article e0243514.
- Silva, V., Santos, M., Rosa, M., 2021. Does she deserve it? The influence of gender and meritocracy in reactions to affirmative action legislation. *Sociol. Sci.* 10 article 386.
- Smyth, F.L., Nosek, B.A., 2015. On the gender-science stereotypes held by scientists. *Front. Psychol.* 6 article 415.
- Snijders, T.A.B., Bosker, R.J., 2012. *Multilevel Analysis*, 2nd ed. Sage, London.
- Solga, H., Rusconi, A., 2007. Determinants of and obstacles to dual careers in Germany. *Zeitschrift für Familienforschung* 19, 311–336.
- Solga, H., Rusconi, A., Netz, N., 2023. Professors’ gender biases in assessing applicants for professorships. *Eur. Sociol. Rev.* 39, 841–861.
- Statistisches Bundesamt, 2020. Personal an Hochschulen, Fachserie 11, Reihe 4.4. Wiesbaden.
- Steinpreis, R.E., Anders, K.A., Ritzke, D., 1999. The impact of gender on the review of the curricula vitae of job applicants and tenure candidates. *Sex Roles* 41, 509–528.
- Su, D., Steiner, P.M., 2020. An evaluation of experimental designs for constructing vignette sets in factorial surveys. *Sociol. Methods Res.* 49, 455–497.
- Urry, M., 2015. Science and gender: scientists must work harder on equality. *Nature* 528, 471–473.
- Vohlidalová, M., 2014. Academic mobility in the context of linked lives. *Hum. Aff.* 24, 89–102.
- Walzenbach, S., 2019. Hiding sensitive topics by design? An experiment on the reduction of social desirability Bias in factorial surveys. *Survey Research Methods* 13, 103–121.
- Williams, W.M., Ceci, S.J., 2015. STEM faculty prefer hiring women professors 2:1. *PNAS Proceedings of the National Academy of Sciences* 112, 5360–5365.
- Wu, M.-J., Zhao, K., Fils-Aime, F., 2022. Response rates of online surveys in published research. *Computers in Human Behavior Reports* 7 article 100206.