GENDER DIFFERENCES IN CONFIDENCE AND ANTICIPATED SCRUTINY IN ECONOMICS PEER-REVIEW*

Margaret Samahita† Martina Zanella‡

October 22, 2026

This paper examines gender differences in stated confidence in a novel, consequential setting: academic conference peer review. Unlike previous studies, female reviewers are at least as confident as male reviewers on average, but a gender difference exists in how confidence relates to expertise. Women declare relatively higher confidence when reviewers' expertise is more uncertain and therefore more likely to be scrutinized. Moreover, organizers' decisions align more closely with male reviewers, although female reviewers' evaluations better predict subsequent citations. These findings suggest stated confidence reflects not only self-assessed expertise but also responses to how that expertise is likely to be evaluated.

*This paper was previously circulated as *"Confident, but Undervalued: Evidence from the Irish Economic Association Conference"*. We thank Manvi Jindal, Hannah Ortega McCormack, and Alban Sims for excellent research assistance, the Irish Economic Association conference organizers 2016–23 for generously sharing their submissions data, and seminar and conference participants for helpful suggestions. This work was supported by the TCD PhD Research Assistantship Programme and TCD Summer Internship Programme for Third Year Economics Students, and the Irish Economic Association. This paper uses confidential data from the Irish Economics Association Conference, which can be obtained by filing a request directly with the Irish Economics Association (`https://www.iea.ie/contact-the-iea/`). The authors hereby declare that they have nothing to disclose and they have no relevant or material financial interests that relate to the research described in the paper.

†Department of Economics and Geary Institute for Public Policy, University College Dublin; email: margaret.samahita@ucd.ie

‡Department of Economics, Trinity College Dublin, TRiSS, and TIME; email: zanellam@tcd.ie.

I Introduction

A large literature documents a gender gap in confidence: across a wide range of domains, women tend to report lower confidence than men, particularly in male-dominated environments (e.g. Jakobsson, 2012; Gordon and Dahl, 2013; Sarsons and Xu, 2021; Sievertsen and Smith, 2025b; Coffman et al., 2024; Exley and Kessler, 2022; Exley and Nielsen, 2024, though see recent review by Bandiera et al., 2022). This confidence gap has traditionally been cited as a key factor behind the undervaluation of female expertise. When women express lower confidence in their own abilities or judgments, their expertise may be discounted—not because it is less valid, but because it is less assertively signaled. As a result, women's lack of confidence has emerged as an important explanation for why observed gender disparities in professional recognition are perpetuated, even in the absence of overt discrimination, differences in opportunities, or constraints.

This paper challenges the prevailing narrative of women's underconfidence. We study stated confidence in a novel and consequential professional evaluation setting: peer review of academic conference submissions. Specifically, we ask whether reviewers' declared confidence in their own expertise varies by gender and context, and how this relates to paper acceptance decisions and subsequent paper citations.

Our setting is interesting for several reasons. As in other professional contexts, expertise is imperfectly observed, so decisions partly rely on signals thereof—making confidence expression central to understanding inequality in professional recognition. Additionally, because perceptions of reviewer credibility may themselves vary by gender, male and female reviewers may differ in how confidently they present their expertise. Yet peer review differs from other settings in an important way. Reviewers express confidence in evaluating *someone else's* work rather than their own. In this setting, the reputational and identity costs of assertiveness may be lower for women, since self-promotion, but perhaps not confidence in evaluating others, may conflict with prevailing gender norms.

We study organizers' and reviewers' decisions at the Irish Economic Association (IEA) Conference, the flagship annual event of the IEA, from 2017–2023. Our empirical strategy exploits a distinctive feature of the review process: for each submission, reviewers rate not only the paper itself but also their own confidence in their expertise. To our knowledge, no other economics conference systematically collects this information, offering a rare opportunity to study confidence in a professional evaluation setting. We supplement these data with information on reviewers, authors, and papers, collected from CVs, institutional/personal webpages, and Google Scholar.

Contrary to the broader literature, which consistently documents lower confidence among women in male-dominated professional settings, we find that female reviewers are *at least as confident* as their male counterparts. Female reviewers report on average 5% higher confidence

than male reviewers when reviewing similar papers and controlling for individual reviewer characteristics. This difference masks substantial heterogeneity. The positive gender gap is driven by professors or equivalent reviewers. For other ranks, women are as confident as men, but among professors and equivalent reviewers, the gap is significantly larger.

What can account for this surprising result? A natural starting point is selection into the profession. If women face barriers to entering and progressing in male-dominated fields, those who make it are likely to be positively selected on characteristics correlated with confidence—such as expertise and experience. Such selection may become increasingly pronounced at higher ranks. We would therefore expect women to report confidence close to men's, or even exceeding men's with seniority.

When we directly examine whether greater experience or expertise can account for female professors' higher confidence, we find that selection into the profession provides, at best, a partial account. The relationship between expertise and confidence differs by gender. The positive gender gap is driven entirely by papers outside reviewers' field of expertise, rather than papers in areas where reviewers have greater topic-specific expertise. Moreover, while male professors express weakly higher confidence when evaluating papers within their field, female professors report significantly higher confidence when evaluating papers *outside* their field than within. The positive gender gap outside reviewers' fields of expertise persists across all levels of both reviewing and professional experience, making it difficult to reconcile with the simple explanation that women are more expert or experienced.

The gender difference in how confidence relates to expertise is particularly pronounced in contexts where women's expertise is likely to be perceived as more uncertain: in these contexts, the gender confidence gap becomes systematically more positive. This contrasts with prior work which finds that women report lower confidence than men when operating outside their field of expertise in male-dominated professions (Sarsons and Xu, 2021; Sievertsen and Smith, 2025b), in stereotypically male domains (Beyer, 1990; Lundeberg et al., 1994; Jakobsson, 2012; Coffman, 2014; Exley and Kessler, 2022), and particularly so when gender identity is more salient (Bordalo et al., 2019). In these contexts, women's expertise may be perceived as more uncertain, increasing reliance on priors and stereotypes in forming expectations about expertise (e.g. Phelps, 1972; Ashenfelter and Albert Rees, 1973; Reuben et al., 2014; Bohren et al., 2019; Coffman et al., 2021). In our setting, however, we find the opposite pattern: women report relatively higher confidence when evaluating papers outside their field of expertise, papers authored by men, and papers in stereotypically male fields. This contrast suggests that the relationship between uncertainty and stated confidence depends on the setting in which confidence is expressed. Unlike much of the previous literature, reviewers in our setting express confidence in evaluating someone else's work rather than their own performance or opinions, potentially reducing the reputational and identity costs of

expressing high confidence. This interpretation is consistent with evidence that the gender gap in negotiation narrows when individuals negotiate on behalf of others (Mazei et al., 2015).

Why might greater uncertainty be associated with higher confidence for women in our setting? One possibility is that this pattern reflects, at least in part, a response to expectations of heightened scrutiny of women's expertise. Given extensive evidence that female economists' expertise is discounted (e.g. Dupas et al., 2025; Sarsons et al., 2021; Boring, 2017; Mengel et al., 2019), women may anticipate greater scrutiny from organizers, particularly when their expertise is less readily observable—an expectation that may be stronger among female professors, who have spent longer in the profession. Such expectations could affect confidence through different channels. On the one hand, anticipating greater scrutiny may induce women to exert more effort, which could in turn increase their confidence. On the other hand, this confidence may reflect considerations beyond a genuine self-assessment of expertise. For instance, such scrutiny may create a stronger incentive to signal competence, leading women to express greater confidence without necessarily exerting greater effort.[1]

We conduct several exercises to examine whether the pattern is consistent with the above interpretation. First, we shift from paper-level to reviewer-level variation, considering two settings in which uncertainty about a reviewer's *overall* competence is greater, independently of gender stereotypes about the paper's content. We consider reviewers from lower-ranked PhD institutions, whose credentials offer less immediate evidence of competence, and reviewers less familiar to organizers, about whom organizers likely have less information. In these contexts, statistical discrimination theory suggests that organizers have greater scope to rely on priors about competence when weighting reviewers' suggestions in their final acceptance decisions. If greater uncertainty about their expertise leads female reviewers to anticipate greater scrutiny, both mechanisms discussed above—greater effort and stronger incentives to express confidence—predict higher declared confidence by women, particularly for those who come from lower-ranked PhD institutions or have lower familiarity with the organizers. This prediction is particularly relevant for women reviewing outside their field, where uncertainty about competence may be more pronounced. We therefore examine whether the relationship between topic expertise and stated confidence differs by gender across these settings. Consistent with this interpretation, the gender difference in how confidence relates to topic expertise is more pronounced among reviewers from lower-ranked institutions and those who are not colleagues of the organizers, translating into a more *positive* gender confidence gap in these settings.

Second, we analyze the comments reviewers leave for organizers in order to test whether higher

[1]Another possible effort-based mechanism is that situations in which women's expertise is more likely to be scrutinized may also increase women's self-doubt (Sarsons and Xu, 2021; Bordalo et al., 2019). This self-doubt could induce women to exert greater effort, which may in turn increase their confidence. Our analysis of reviewing effort provides little support for this channel, as discussed in Section VC.

confidence is associated with greater reviewing effort. We find that commenting behavior follows the opposite pattern to confidence: men leave more comments, in particular when reviewing outside their field.[2] Assuming that more diligent reviewers are more likely to leave feedback, these results provide little support for effort as a channel through which anticipated scrutiny raises women's confidence. While our data do not allow us to identify the specific mechanism, the findings are suggestive of an explanation in which stated confidence reflects, at least for women, considerations beyond a genuine self-assessment of expertise rather than greater reviewing effort.

If women did expect higher scrutiny, these expectations would be well-founded. We document an influence gap: controlling for individual reviewer characteristics, female reviewers' recommendations are less predictive of organizers' decisions on similar papers than male reviewers' recommendations. On average, this gap arises for marginal submissions, where organizers have the greatest discretion. Among more senior and more confident reviewers, however, the gap extends to papers they recommend accepting. Together, these patterns suggest that women's recommendations receive less weight in organizers' decisions. Although this does not directly reveal women's beliefs, the evidence of differential weighting is consistent with the possibility that women anticipate greater scrutiny of their expertise.

Finally, we examine whether the differential weighting of male and female reviewers' evaluations is associated with differences in decision quality. We compare the number of citations accumulated in the years following the conference for papers evaluated by reviewers of different genders that received the same evaluation score, controlling for reviewer and paper characteristics. Female reviewers' evaluations are more strongly correlated with papers' subsequent citations. Thus, female reviewers' assessments are, if anything, more predictive of paper quality—despite being less influential in the conference organizers' decision.

Taken together, the evidence provided by this paper suggests that when individuals provide professional evaluations of colleagues' work under imperfectly observed expertise—and may fear that their credibility will be more strongly questioned—they may respond in ways that increase their observable confidence. While our analysis focuses on women in economics, the underlying features of our setting are common to many other environments where underrepresented groups evaluate others' work. Similar dynamics may arise in law, where judges review colleagues' opinions (Ash et al., 2024), or in medicine, where general practitioners assess specialists' recommendations (Sarsons, 2017)—and more broadly in any professional context where evaluators' expertise is not fully observable, and their assessments are interpreted alongside signals like confidence. Our results, therefore, speak to a broader class of settings characterized by high discretion, imperfectly observed expertise, and the potential for evaluators from particular groups to anticipate greater

[2]We consider both the probability of leaving a comment and an effort score created by an LLM asked to classify comments based on the level of effort put into the review. Both measures yield similar results.

scrutiny of their competence.

Our paper contributes to several strands of the literature. First, it enriches our understanding of the existence, magnitude, and drivers of gender differences in confidence. A large body of work documents that women exhibit lower levels of confidence (Jakobsson, 2012; Barber and Odean, 2001; Buser et al., 2014), especially in male-dominated domains (Lundeberg et al., 1994; Beyer, 1990; Möbius et al., 2022; Bordalo et al., 2019; Coffman et al., 2024; Exley and Kessler, 2022; Coffman, 2014).[3] Similar patterns arise when individuals are asked to publicly evaluate issues or offer opinions (Gordon and Dahl, 2013; Sarsons and Xu, 2021; Sievertsen and Smith, 2025b). Our findings qualify this view by showing that gender differences in stated confidence are highly context-dependent. Unlike most existing work, which focuses on self-assessments or personal opinions, we study confidence in the evaluation of *others'* work—a setting where identity costs likely differ, and incentives may exist to challenge the societal expectation prescribing women to be less assertive. In our setting, we find that women express higher confidence compared to men precisely where the existing literature would have predicted otherwise.

Second, our paper relates to a broad literature documenting that women's expertise is often discounted relative to men's, across settings including economics (e.g. Doleac et al., 2021; Hospido and Sanz, 2021; Koffi, 2021; Boring and Philippe, 2021; Lundberg and Stearns, 2019; Baltrunaite et al., 2025; Eberhardt et al., 2023; Card et al., 2020; Hengel and Moon, 2020; Hengel, 2022, as well as the studies cited above), medicine (de Vaan and Stuart, 2022; Sarsons, 2017), finance (Klein et al., 2021), and hiring (Radbruch and Schiprowski, 2023). We contribute to this literature in two ways. First, we provide novel evidence that this pattern extends to peer review, where women's recommendations receive less weight than men's. Second, we provide suggestive evidence of a previously unexamined "feedback loop" between gender differences in confidence and disparities in professional recognition: when expertise is imperfectly observed, expectations of greater scrutiny may shape how confidence is expressed.

Third, we contribute to the emerging literature on responses to anticipated discrimination. Whether individuals actively adopt strategies to mitigate the adverse consequences of discrimination remains understudied, as such responses are inherently hard to measure. Prior work has documented identity concealment (Lane, 2024; Kang et al., 2016; Charness et al., 2020; Arai and Skogman Thoursie, 2009; Biavaschi et al., 2017; He et al., 2024; Zussman, 2013; Abraham et al., 2023; Aksoy et al., 2023; Kudashvili and Lergetporer, 2022) and avoidance behaviors (Lepage et al., 2025; Ulgade, 2022; Reubeck, 2025; Angeli et al., 2026). In economics, for example, the EEA professional climate survey suggests that anticipated discrimination leads women to avoid seminars, conferences, social events, and even entire research fields (Lee et al., 2025). A few

[3] Although a recent paper by Bandiera et al. (2022) casts doubt on some of the conclusions that can be drawn from the existing evidence.

papers document a different response: potentially costly investments in signaling, consistent with anticipated discrimination in credential evaluation at hiring (Exley et al., 2024; Batistich et al., 2024; Lang and Manove, 2011).[4] We point to another response: anticipating heightened scrutiny of their expertise may shape how individuals express confidence in their professional judgments. Our setting provides a rare opportunity to study this possibility in a natural professional environment with consequential tasks, real reputational stakes, and limited scope for experimenter demand effects.

The remainder of the paper is organized as follows. Section II provides a background on the IEA Annual Conference structure and an overview of the data. Section III presents the baseline evidence on gender differences in stated confidence. Sections IV and V examine how these differences vary with reviewers' expertise, experience, and the context of the evaluation. Section VI discusses the evidence on the gender influence gap and its implications for paper outcomes. Finally, Section VII concludes.

II Background and Data

The IEA Annual Conference The IEA Annual Conference is the flagship event of the Irish Economic Association. It takes place every year in May in a different institution across Ireland. Around 150 papers are submitted every year by both national and international economists from academic and other research institutions. Around 60% of the papers come from one of the academic and non-academic institutions on the island, with the remaining 40% coming from institutions outside Ireland (Table A.1). Submission closes in February, after which the conference organizers assign papers to a pool of around fifty reviewers according to their field of research. Around 60% of submitted papers are accepted every year.

The review process Reviewers include economists from both Irish and international institutions. Most reviewers are drawn from a pool of reviewers who participated in previous editions of the conference. The list is updated each year to include new experts as needed, ensuring sufficient coverage across research areas so that submitted papers can be matched with appropriate reviewers and that no reviewer reviews more than two to three papers.

Papers are allocated according to the closest match between the paper's topic and the reviewer's field of expertise, following a process similar to that used by academic journals. Based on discus-

[4]Exley et al. (2024) find that female university students are less likely than males to conceal GPA-harming grades, consistent with expectations that employers will draw worse inferences from concealed grades. Batistich et al. (2024) and Lang and Manove (2011) provide evidence that Black students select more difficult, higher-paying majors than white students with similar academic preparation, consistent with signaling ability to employers in anticipation of statistical discrimination in hiring.

sions with conference organizers, assignments are made using organizers' knowledge of reviewers' research areas.[5] Reviewers receive an invitation to review and may decline if they do not feel sufficiently qualified to assess the submission, or for other reasons. When no suitable reviewers could be found, or when invited reviewers fail to submit reports, conference organizers or their colleagues undertake the reviews themselves.[6][7]

Each paper is assigned to a single reviewer.[8] Reviewers have one month to complete their evaluations. They assess papers based on technical merit, readability, originality, and relevance to the conference, using a scale from 1 (unacceptable) to 5 (excellent). Reviewers also provide an overall **acceptance score**: 1 (do not accept), 2 (weak reject), 3 (borderline reject/accept), 4 (probably accept), and 5 (definitely accept). Along with their ratings, reviewers can submit a brief comment to the organizers. Additionally, reviewers are asked to provide their **confidence score**. The question asked is "As a reviewer, how confident were you within the knowledge area discussed in this submission?", which captures the reviewer's perceived expertise in the field. Reviewers can select one of the following options: "1 No Confidence: I am not qualified to pass judgment on this submission," "2 Low Confidence: I do not have enough experience to make a definitive decision on this submission," "3 Some Confidence: I have a reasonable understanding of this research area," "4 Confident: I have considerable experience in this subject area," or "5 Very Confident: I am an expert in this field of research." To the best of our knowledge, the IEA Conference is unique among economics conferences in systematically collecting this information. Screenshots of the webpage that reviewers are presented with can be found in Appendix Figures B.1 and B.2.

In 2021, and 2022, the IEA implemented a blind submission system that concealed author names from reviewers. However, reviewers who were also conference organizers could still see all author names. For our analyses, such submissions will be classified as unblinded, along with those from 2017-2019 and 2023.[9]

The dataset We use data from the 2017–2023 conferences.[10] The data comes from the conference organizers, who manage submissions through the centralized Ex Ordo system. The dataset consists of 900 observations, each representing a single paper submission. It includes details provided by

[5] Although this information is not directly observable to us, we systematically collect information on each reviewer's research fields through a manual review of online profiles, as described below.

[6] We can only observe the final reviewer of each paper and cannot determine whether it was initially assigned to someone else.

[7] Information on the resulting reviewer–submission matches can be found in Table 1 and is discussed in the "Descriptive statistics" section.

[8] Our setting, therefore, does not allow the comparison of evaluations across different reviewers for the same paper.

[9] Some authors did not follow the instruction to anonymise and included their names in the PDF submitted. Each case was reviewed individually, and only fully anonymized submissions were classified as blinded, while the rest were classified as unblinded.

[10] The 2020 conference was canceled due to the Covid-19 pandemic. We also have data from 2016, but no confidence question was asked that year.

the submitter, such as the paper's title, abstract, primary field, and optional secondary field,[11] as well as whether it is a student paper. It also contains submitter information, including title, full name, email, institution, and country for both the submitter and co-authors. Additionally, the dataset records reviewer evaluations, including their acceptance score and confidence score, as well as whether the paper was eventually accepted by the organizers.

We supplement the conference dataset with additional information on the papers, their submitters and co-authors, and the reviewers. Through a manual search of personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar, we collected data on individuals' gender, research fields, PhD status, PhD graduation year, affiliation, position, number of solo- and co-authored publications, and total citations up to December 31st of the year preceding the conference. To assess paper quality, we collected yearly citation counts for each submission and whether the paper was published as of January 2025, including details on the journal.

The sample of analysis consists of 853 submissions with non-missing reviews across 2017-2023 (~ 150 per year).[12]

Descriptive statistics The sample of reviewers consists of 132 unique reviewers over the analysis period, primarily from Irish academic and non-academic institutions. Reviewer characteristics are shown in Tables A.2 and A.3, presented at the reviewer-year level to account for reviewers moving institutions or being promoted. Half of the reviewers obtained their Ph.D. from a university outside Ireland (49.9%), mostly in other European countries or in the US. 46.7% of reviewers hold a Ph.D. from an Irish university—most of whom earned their degree from Trinity College Dublin or University College Dublin. Just under half of the reviewers are junior economists, with the largest group (34.5%) being Assistant Professors or equivalent. The remaining 55.6% of reviewers are more senior economists, half of whom (27.80%) are Professors or equivalent. On average, reviewers have 16 years of post-Ph.D. experience, 1120 citations, and 21 published papers at the time of the conference. Most have prior experience reviewing for the conference, with an average of 6 previously reviewed papers. Finally, 28% of reviewers are female.

Table 1 presents descriptive statistics on paper allocation across reviewers. Female and male reviewers do not review systematically different papers. On average, roughly 40% of reviewer–paper assignments represent a perfect match within the same research field. The fact that many papers are not perfectly matched likely reflects several factors: the limited granularity of research area

[11] Submitter selects the paper's field out of the following: Agricultural and Natural Resource Economics, Economic Development and Growth, Economic History, Economic Systems, Energy Economics, Financial Economics, Health, Education, and Welfare Economics, History of Economic Thought, Household Finance and Consumption, Housing Economics, Industrial Organisation, International Economics, Labour/Demographic Economics, Macroeconomics, Microeconomics, Public Economics, Regional/Real Estate/Transport Economics, Teaching Economics.

[12] 47 papers have missing reviews. These papers were evaluated directly by the organizers when reviewers failed to submit their reviews by the deadline. The organizers made acceptance or rejection decisions without assigning an acceptance score.

categories that submitters can select in Ex Ordo (see footnote 11), the interdisciplinary nature of submissions, and practical constraints such as balancing reviewer workload with subject-matter fit. Crucially, this allocation pattern does not differ significantly by reviewer gender. Male and female reviewers both review on average 72% of papers by male authors. Similarly, the number of authors or the proportion of Irish economists authoring the paper is consistent across genders. Very few papers had citations before the conference year, and around 27% of papers are blinded; these characteristics are also balanced across genders. However, male reviewers evaluate significantly more papers from authors with a PhD than female reviewers. Female reviewers also review a slightly higher proportion of papers by authors from the same institution, though the difference is not significant at the 5% level.[13] Overall, this shows a fairly balanced paper assignment across reviewers of different genders.

Table 1- Paper Allocation

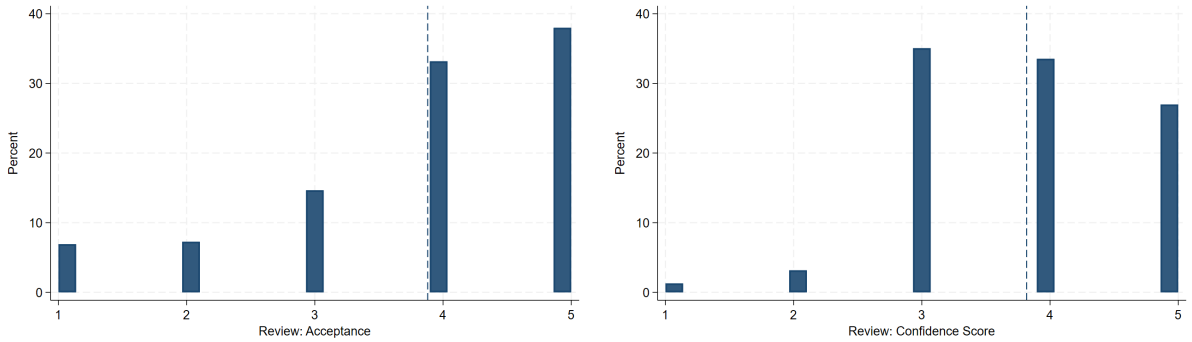
	Male Reviewers		**Female Reviewers**		**T-test**
	N.	Mean	N.	Mean	Diff. p-value
Within field of expertise	622	0.42	231	0.37	0.15
Within field of expertise (not organizers)	503	0.45	187	0.42	0.47
Blinded	622	0.21	231	0.24	0.32
Blinded (not organizers)	503	0.26	187	0.30	0.31
Male submitter	622	0.71	231	0.69	0.62
Prop. of male authors	622	0.72	231	0.72	0.88
Prop. of authors with PhD	622	0.74	231	0.66	0.01
Prop. of authors affiliated with Irish inst.	622	0.55	231	0.53	0.57
Prop. of authors affiliated with Irish inst. with WP series	622	0.39	231	0.38	0.85
Number of authors	622	2.05	231	2.10	0.62
Authors same institution	622	0.05	231	0.08	0.06
Citations (pre-conference)	622	0.33	231	0.35	0.92

Notes: Descriptive statistics on the full sample of submissions to the IEA Conference from 2017-2023. Within field of expertise is a dummy variable which equals 1 if the paper's primary field, as stated by the submitter, matches one of the reviewer's fields, based on data from personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. N. citations (pre-conference) is imputed to 0 if we couldn't find information on the number of citations for the paper. Blinded is a dummy equal to 1 if submissions were fully anonymized.

Figure 1 shows the distribution of reviewer scores. The average acceptance score is 3.88, with responses distributed as follows: 38% "definitely accept," 33% "probably accept," 15% undecided, and 14% either "weakly reject" or "do not accept." Reviewers report high confidence overall, with an average score of 3.82 (SD = 0.91). Most reviewers (60%) are either confident (34%) or very confident (27%), while 35% are somewhat confident and 4.50% have low or no confidence.

[13] If we exclude organizers, the proportion of authors from the same institution is also balanced across genders.

Figure 1- Distribution of reviewers' acceptance scores and confidence scores



(a) Acceptance score

(b) Confidence score

Notes: Histogram of reviewers' acceptance scores (panel a) and confidence scores (panel b). The vertical dashed lines indicate the average scores.

III Gender Differences in Stated Confidence

We begin by examining gender differences in stated confidence. We compare the confidence scores provided by reviewers of different genders, controlling for an array of paper and reviewer characteristics. We estimate:

$$Y_{ipt} = \alpha_t + \beta Female_i + X_i'\gamma + X_p'\delta + \epsilon_{ipt} \tag{1}$$

where Y_{ipt} denotes the score reviewer i gave to paper p in year t. We include a set of reviewer-specific controls X_i': affiliation group FEs, N. years post Ph.D. (second order polynomial), Ph.D. institution group FEs, rank FEs, N. of publications (solo and co-authored separately), N. of citations at the time of the conference (second order polynomial), and a dummy for papers in the field of expertise. Moreover, we control for a range of paper-specific characteristics X_p': primary field, a blinding dummy, N. authors on the paper, the proportion of authors who are: male, with Ph.D., affiliated with an Irish institution, affiliated with an Irish institution with a Working Paper (WP) series, a dummy equal to one if submitter's institution is the same as the reviewer's, and N. citations of the paper before the conference.[14] We include year FEs, denoted α_t.[15]

We find an average positive gender gap. Female reviewers report an average confidence score of 3.99 (on a 1-5 scale), compared with 3.75 for male reviewers (t-test, $p = 0.0009$), as shown in Figure A.1. This difference remains when controlling for the characteristics of the paper and

[14]The conference application deadline is typically in February. The paper's number of citations refers to the number of citations that the working paper (if available) had by the end of the year preceding the conference.

[15]We do not control for the acceptance score assigned by the reviewer, as it is jointly determined with confidence.

the reviewer: on average, women report confidence scores 0.18 points higher than men when evaluating papers with similar characteristics and holding reviewer characteristics constant (Table 2). This corresponds to 5% of the average confidence score of 3.82, or 0.20 standard deviations. Full regression results, including all covariates, are presented in Table A.4.

Table 2- Gender gap in declared confidence

	Reviewers' confidence score {1;5}				
	(1)	(2)	(3)	(4)	(5)
Female	0.19***	0.22***	0.19**	0.18**	0.21**
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
N	853	853	853	853	699
Year FE	X	X	X	X	X
Paper controls		X	X	X	X
Referee controls			X	X	X
Rank fixed effects				X	X
Complete information					X

Note: This table reports the estimated gender gap in confidence. The regression specification is given by equation (1). Ph.D. institution group FEs and rank FEs include a group for missing observations. The paper is labeled as outside the reviewer's field of expertise if no information on the field of interest has been found for the reviewer (6 observations). The number of publications, citations, and years post-PhD are imputed to the average if we could not find this information (31, 1, 30 reviewer-year pairs, respectively). We include a dummy variable indicating whether the observation has an imputed value. Column (5) includes only the reviewer-paper pairs for which we have a complete set of information. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The finding that female reviewers do *not* report lower confidence than their male counterparts stands in contrast to the broader literature, which consistently documents lower confidence among women in male-dominated professional and academic settings (Sarsons and Xu, 2021; Jakobsson, 2012; Barber and Odean, 2001; Lundeberg et al., 1994). This warrants further exploration. Indeed, the average gender gap may mask important variation across reviewers or the circumstances in which they evaluate papers. In the sections that follow, we examine potential sources of this heterogeneity and use these patterns to shed light on the underlying drivers of the gender confidence gap.

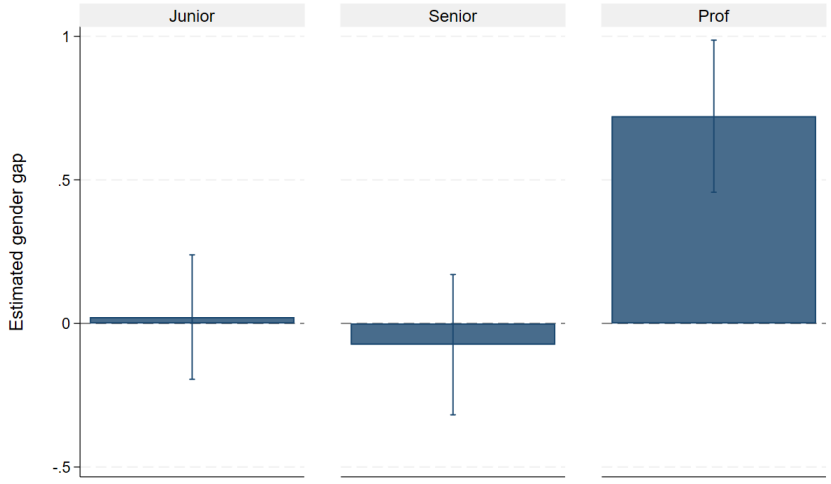
IV Variation across Reviewers: Selection into the Profession

A natural starting point is to ask whether selection into the profession can account for the positive gender gap. If women must be more like men to succeed in male-dominated fields, we might expect gender differences in confidence to disappear in these settings. They might even reverse at higher levels of seniority (Adams and Funk, 2012), if women who reach the top have had to overcome greater barriers to entry and are therefore positively selected on expertise or other characteristics correlated with confidence. We therefore examine whether the gender difference in confidence

varies with seniority in the profession.

The results obtained from estimating equation (1), including interactions between the reviewer's gender and rank, provide some evidence consistent with this possibility. Figure 2 shows that the gender difference in confidence is not significantly negative at any rank, but is significantly positive among professors and equivalent reviewers.[16]

Figure 2- The gender confidence gap across ranks



Notes: Bar graph that displays the estimated gender confidence gap (female minus male) across ranks, net of reviewer and paper characteristics as in equation (1). Junior includes Assistant Professor or equivalent, or below. Senior is a dummy equal to one if the reviewer is an Associate Professor or equivalent. Prof is a dummy equal to one if the reviewer is a Professor or equivalent. The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full regressions are shown in column 3 of Table A.5.

Can selection into seniority account for this pattern? While women can be selected along various characteristics, our data enable us to examine whether two such characteristics that are potentially related to confidence, namely expertise and experience, can explain the observed heterogeneity.

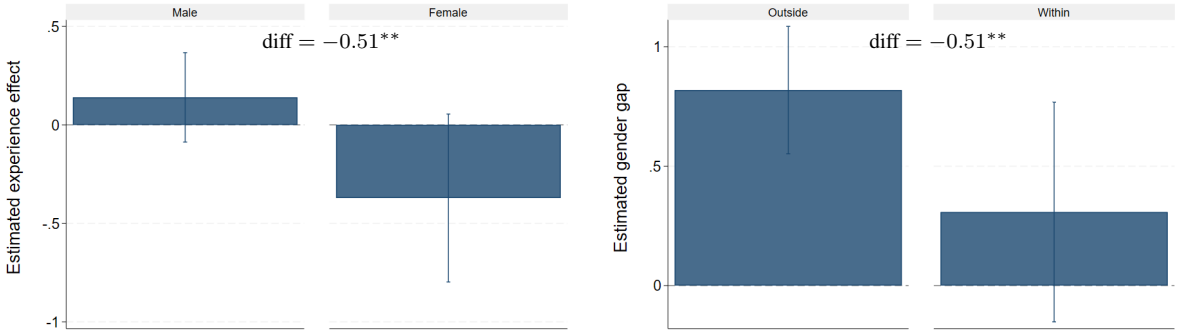
First, we examine whether differences in expertise are responsible for this pattern by assessing whether confidence differs between papers within and outside the reviewer's field of expertise. We define a paper as within the reviewer's field if its primary field, as declared by the author, matches one of the reviewer's stated fields of interest (from personal and institutional webpages, CVs,

[16]One reviewer accounts for 54% of observations involving female professors. Excluding her makes the professor subsample smaller, but female professors still report significantly higher confidence than male professors, albeit with a smaller gap (0.41, $p = 0.03$, vs. 0.72, $p < 0.01$, at baseline). As a result, the average gender confidence gap shrinks and loses significance when we exclude this reviewer. Our subsequent results, showing that male and female reviewers' confidence responds differently to topic expertise across contexts, are robust to her exclusion. Full results are in Appendix D.

LinkedIn profiles, and Google Scholar).[17] We estimate specification (1) including interactions between this expertise measure and the female and professor indicators to estimate the effect of expertise on confidence by gender and rank.

The results provide little support for an explanation based on positive selection on expertise. Figure 3, panel b, shows that female professors are not significantly more confident than male professors when reviewing papers within their field, while the estimated gender difference is substantially larger for papers outside their field (0.31, $p = 0.19$ versus 0.82, $p < 0.001$, difference$= 0.31 - 0.82 = -0.51$, $p = 0.04$). Thus, the positive gender gap estimated for professors is concentrated in evaluations of papers outside the reviewer's field rather than in areas where reviewers have greater topic-specific expertise.

Figure 3- The role of expertise for professors



(a) Estimated expertise effect by reviewer gender **(b)** Estimated gender gap by topic expertise

Notes: Bar graph that displays the estimated effect of being an expert on the paper topic (within minus outside the field of expertise) for male and female professors (panel a) and the estimated gender confidence gap for professors (female minus male) for papers outside and within the field of expertise (panel b), net of paper and reviewer controls as specified in equation (1). Within expertise is a dummy equal to 1 if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full regressions are shown in Table A.6.

A broader interpretation of positive selection could nevertheless generate this pattern. If all professors are highly expert within their own fields, differences in expertise between male and female professors may be difficult to detect for papers within the field. Moreover, if female professors are more broadly knowledgeable, their greater expertise could manifest itself as higher confidence when reviewing papers outside their specific field. However, even under this interpretation, we

[17]Table 1 shows that the share of papers within and outside the field of expertise allocated to reviewers does not systematically vary by gender. In Appendix C we provide a different measure of field expertise using LLM to categorize reviewers' work; our results are robust to this alternative definition.

would expect female professors to be at least as confident when reviewing papers *within* their field as outside it, since their expertise should be greatest in their own field.

We find the opposite pattern. Figure 3, panel a, shows that male professors are slightly more confident when reviewing papers within their field than outside it (0.14, $p = 0.23$), whereas female professors report *lower* confidence within their field than outside it (–0.37, $p = 0.09$). Greater topic expertise is thus not associated with higher confidence among female professors. The larger gender gap outside the field instead points to a systematic gender difference in how confidence relates to expertise.[18]

A similar pattern emerges when we examine two measures of experience – tenure in the profession and reviewing experience. While the gender difference in confidence increases with the number of years spent in the profession after the PhD, women display higher confidence at every level of experience when reviewing papers outside their field of expertise (Appendix Figure A.2). Reviewing experience provides a similar insight. When we estimate the gender difference in confidence for reviewers with different levels of reviewing experience, proxied by the number of papers reviewed in previous IEA conferences, even rookie reviewers—those with no prior reviewing experience—exhibit a positive gender difference in confidence when evaluating papers outside their field of expertise. Moreover, the effect of topic expertise is again negative for women's confidence and positive for men's (Appendix Table A.7).[19]

The evidence therefore suggests that while selection into the profession may contribute to differences in reported confidence, it cannot fully account for the broader pattern. Instead, the gender gap in confidence appears closely tied to a gender difference in how confidence relates to expertise, which persists across levels of seniority and reviewing experience, pointing to factors beyond selection as potential explanations.

V Context: When Do Women Report Higher Confidence?

The previous section shows that the gender difference in confidence is not simply explained by selection into the profession or by differences in experience. Instead, it reflects systematic differences in how expertise relates to confidence across genders. This raises the question of when these differences emerge. We therefore examine how the gender gap in confidence varies across

[18]This pattern is not specific to senior women. Appendix Table A.6 shows that the effect of expertise on confidence is negative for women and positive for men among non-professors as well, with the gender difference statistically significant at both ranks (Professors = –0.51**, Non-Professors = –0.45***). We return to this finding in the next section.

[19]Experience in reviewing papers is strongly correlated with declared confidence (right panel of Figure A.3). Moreover, the rank in the profession is positively correlated with reviewing experience, more so for women than for men (left panel of Figure A.3).

contexts, identifying the settings in which women report higher confidence and what these patterns reveal about its drivers.

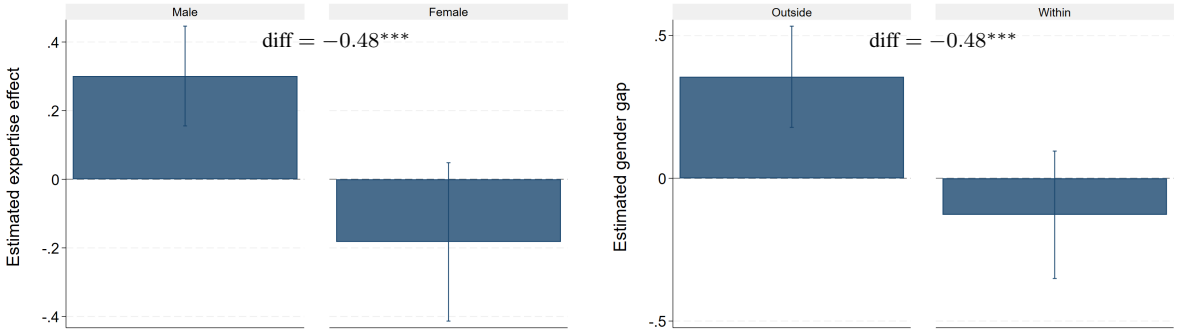
VA Confidence under Uncertainty

We start with three settings motivated by the literature on gender differences in confidence: whether the paper lies within the reviewer's field of expertise, the gender of the author, and the gender stereotypicality of the topic. Prior work finds lower confidence among women in these contexts. At the same time, these settings may involve greater uncertainty about the reviewer's expertise, either for the reviewer themselves or for others evaluating their competence, and greater scope for gender stereotypes to shape these evaluations.

Field of expertise We first revisit the distinction between papers within and outside the reviewer's field of expertise introduced in Section IV. We estimate specification (1) for all reviewers, including an interaction between the expertise dummy and the female dummy, to examine whether the relationship between expertise and confidence differs by gender. Figure 4 displays the results.

The pattern we observe for professors (Section IV) extends to the full sample. Men report significantly higher confidence when reviewing papers within their field of expertise, in line with them having greater topic-specific expertise. Women do not. The gender difference in the effect of expertise on confidence is substantial and negative (interaction coefficient -0.48, $p < 0.01$; see Appendix Table A.8), indicating a striking difference in how confidence responds to expertise for men and women. Consistent with this difference, a positive gender gap in confidence emerges when reviewers evaluate papers outside their field (panel b), exactly where uncertainty about expertise is greater. There is no significant gender gap for papers within the field of expertise.

Figure 4- The role of expertise



(a) Estimated expertise effect by reviewer gender **(b)** Estimated gender gap by topic expertise

Notes: Bar graph that displays the estimated effect of being an expert on the paper topic (within minus outside the field of expertise) for men and women (panel a) and the estimated gender confidence gap (female minus male) for papers outside and within the field of expertise (panel b), net of paper and reviewer controls as specified in equation (1). Within expertise is a dummy equal to 1 if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full regressions are shown in column 3 of Table A.8.

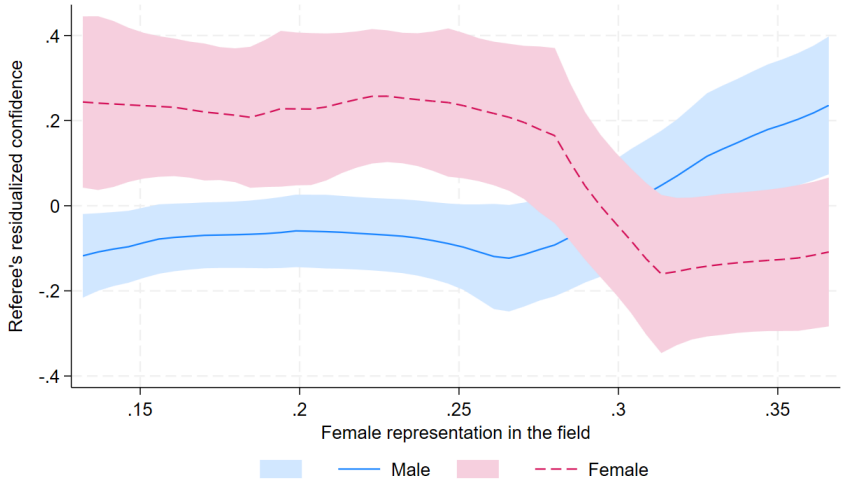
Stereotypically male fields As a second case, we examine whether the gender difference in confidence varies with the gender stereotypicality of the paper's topic. Prior work documents lower confidence among women in male-dominated domains (Beyer, 1990; Lundeberg et al., 1994; Jakobsson, 2012; Coffman, 2014; Exley and Kessler, 2022), which has been attributed to the influence of negative stereotypes about women's competence in these fields. Such domains may also increase uncertainty about how one's expertise will be perceived or evaluated. If the gender difference in confidence reflects women's response to greater uncertainty about their expertise, we would therefore expect it to be larger for papers in stereotypically male fields.

Figure 5 displays the relationship between the gender stereotypicality of the paper's primary field and the confidence score reported by male and female reviewers, net of paper and reviewer characteristics as described in equation (1). We measure a field's gender stereotypicality using the share of women working in it (following, e.g., Zanella, 2024; Bostwick and Weinberg, 2022; Kugler et al., 2021; Hebert, 2025), using statistics from IDEAS. Each paper's primary field is declared by the submitting author, and the resulting categorization is reported in Appendix Table A.9. Formal tests are reported in Appendix Table A.10.

Women report higher confidence when the paper's topic belongs to a more male-dominated field—that is, precisely in the contexts where prior literature finds women to be less confident (Jakobsson, 2012; Exley and Kessler, 2022). A similar, though less pronounced, pattern is observed

among men. As with the field-of-expertise setting, the results are consistent with women stating higher confidence in conditions of greater uncertainty about how their expertise will be perceived or evaluated.

Figure 5- The role of gender stereotype of paper topic



Notes: The graph displays the correlation between the share of women working in the paper's primary field (x axis) and the confidence score provided by male (blue line) and female (red line) reviewers (y axis) for each paper. The red and blue lines display the smoothed values of a kernel local polynomial regression of the residuals from a regression of the reviewers' confidence score on paper and reviewer characteristics, as described in equation (1). The shaded area represents 95% confidence intervals. Female representation in each field is constructed using the statistics regarding the share of women working in each field in Economics provided by IDEAS. The categorization can be found in Table A.9. Formal tests are reported in Appendix Table A.10.

Authors' gender As a third test, we examine whether the gender confidence gap varies with the gender of the paper's author. Prior work documents that women exhibit lower confidence when their gender identity is made more salient, triggering identity-related considerations (Bordalo et al., 2019). In our setting, reviewing a paper authored by a man may similarly make the reviewer's gender more salient, potentially making gender stereotypes about competence more relevant.

The results are presented in Figure 6, which shows the gender confidence gap by the gender of the submitter for unblinded submissions, net of paper and reviewer characteristics as described in equation (1). Appendix Table A.11 presents the corresponding formal tests and reports an alternative definition of male-authored papers based on the proportion of male authors, with qualitatively unchanged results.[20]

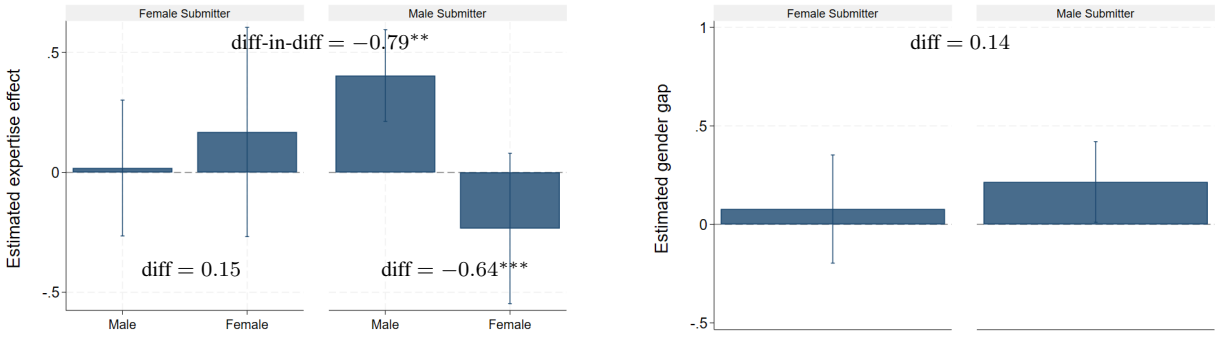
On average, reviewer gender does not interact significantly with submitter gender (column 1, Appendix Table A.11). The interaction becomes significant, however, when we distinguish

[20]Recall that the share of papers with male submitters allocated to reviewers does not systematically vary by gender as shown in Table 1.

between papers within and outside the reviewer's field of expertise (panel a of Figure 6). For papers by female submitters, the effect of topic-specific expertise is positive for both men and women, with no significant gender difference (0.15, $p = 0.57$). For papers by male submitters, however, confidence rises with topic-specific expertise for male reviewers but falls for female reviewers, resulting in a significant gender difference in the effect of expertise (-0.64, $p < 0.01$). The two differences differ significantly from each other ($-0.64 - 0.15 = -0.79$, $p = 0.01$)[21], suggesting that women's confidence responds differently to topic-specific expertise compared to men's when their gender identity is made salient: when reviewing male authors' papers.

These patterns translate into a positive aggregate gender confidence gap when the submitter is male (0.22, $p = 0.04$, versus 0.08, $p = 0.58$; column 1, Appendix Table A.11), as shown in panel b of Figure 6.

Figure 6- The role of the submitter's gender



(a) Estimated expertise effect by submitter and reviewer's gender

(b) Estimated gender gap by submitter gender

Notes: Bar graphs that display the estimated expertise effect on confidence (within minus outside the field of expertise) by submitter and reviewer's gender (panel a) and the estimated gender confidence gap (female minus male) by submitter's gender (panel b), net of paper and reviewers' controls as specified in equation (1). The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full regression results corresponding to panel a and panel b are reported in columns 2 and 1, respectively, of Table A.11.

Taken together, these three exercises show that women's confidence is relatively *higher* where uncertainty about their expertise is likely greater—when reviewing papers outside their field, papers authored by men, and papers in stereotypically male fields. By contrast, men's confidence responds less strongly to these contextual cues or, in the case of field expertise and male-authored papers, in the opposite direction. Women and men, therefore, respond differently to uncertainty about their expertise, resulting, in all of these settings, in a positive gender difference in confidence.

This pattern contrasts with prior work on gender differences in confidence among economists.

[21]This estimate corresponds to the triple-interaction coefficient reported in column 2, Appendix Table A.11.

Sarsons and Xu (2021); Sievertsen and Smith (2025b) find that senior female economists express less confidence than men when responding to questions or voicing opinions outside their field of expertise. Two explanations have been proposed. First, women may more accurately judge the limits of their knowledge and adjust their confidence accordingly when speaking outside their field. Second, women may be more reluctant to voice strong opinions in a field broadly coded as male, such as economics, and this reluctance may be amplified when they are furthest from being perceived as the "legitimate" expert.

Our context shares important similarities with those studied in the previous literature, but also differs in ways that may be consequential for stated confidence. As in Sarsons and Xu (2021); Sievertsen and Smith (2025b), competence is imperfectly observed, the field is male-dominated, and we consider situations in which uncertainty about women's expertise is likely greater. Our setting, however, is more private than those examined in prior work, and reviewers report confidence in evaluating someone else's work rather than in expressing their own opinions. This may entail lower reputational and identity costs than in settings where expressing a strong opinion can invite backlash against women (Rudman and Phelan, 2008; Mazei et al., 2015). The difference in context may help explain why uncertainty is associated with higher rather than lower stated confidence among women in our setting. When the cost of appearing overconfident is high, greater uncertainty may lead women to suppress their confidence. Where this cost is lower, as may be the case in our setting, the same underlying uncertainty may instead prompt women to behave in ways that compensate for this uncertainty, resulting in higher rather than lower stated confidence.

VB Anticipated Scrutiny

One potential explanation for women's higher stated confidence when uncertainty about their expertise is greater is that they anticipate greater scrutiny of their work. Extensive evidence suggests that women's expertise is discounted in the economics profession (e.g. Dupas et al., 2025; Mengel et al., 2019; Boring, 2017; Doleac et al., 2021; Hospido and Sanz, 2021; Koffi, 2021; Boring and Philippe, 2021; Lundberg and Stearns, 2019; Baltrunaite et al., 2025; Eberhardt et al., 2023; Card et al., 2020; Hengel and Moon, 2020; Hengel, 2022). Women may therefore expect greater scrutiny from organizers when reviewing papers in contexts where their expertise is less certain.

Expectations of heightened scrutiny could affect confidence through different channels. Anticipating greater scrutiny may lead women to exert greater effort, which could in turn increase their confidence. Alternatively, stated confidence may reflect considerations beyond a real self-assessment of expertise. For instance, expectations of heightened scrutiny may create stronger incentives to signal competence by reporting higher confidence, without necessarily increasing

effort.[22]

To examine whether the evidence is consistent with this interpretation, we shift from paper-level to reviewer-level variation, considering two characteristics that may affect how readily organizers can assess a reviewer's expertise: the rank of the institution at which they obtained their PhD and whether they are familiar to the organizers. Reviewers from lower-ranked institutions may have less readily observable signals of expertise, while organizers are likely to have less information about reviewers who are not their colleagues. Given that economics is broadly coded as a male field, women may be particularly likely to anticipate scrutiny of their expertise when their competence is less readily observable. Therefore, if this is the case, the mechanisms outlined above predict a more positive gender confidence gap among reviewers from lower-ranked institutions and among those who are less familiar to organizers.

In the following paragraphs, we discuss the difference in confidence scores reported by female and male reviewers, distinguishing between colleagues—those working at the same institution as the organizers—and non-colleagues, and reviewers from lower-ranked versus higher-ranked PhD institutions[23], net of paper and reviewer characteristics as described in equation (1).[24]

Higher- and lower-ranked PhD institutions Institutional rank itself is strongly associated with confidence, but in opposite directions for male and female reviewers. Female reviewers report, on average, 0.55 points ($p = 0.002$) higher confidence when they come from lower-ranked rather than higher-ranked institutions. For male reviewers, the relationship is significantly weaker and runs in the opposite direction: those from higher-ranked institutions report 0.13 points ($p = 0.333$) higher confidence (Appendix Table A.12).

The differential correlation between institutional rank and confidence across genders is driven by reviewers who evaluate papers outside their field of expertise—precisely the context in which organizers have less information with which to assess a reviewer's competence. Figure 7, panel a, shows that among reviewers from higher-ranked institutions, reviewing papers within their field of expertise is associated with higher confidence for both men and women (although not significantly so for women), with no significant gender difference (-0.23, $p = 0.392$). The gender difference in the expertise effect is instead significant for reviewers from lower-ranked institutions (-0.41, $p = 0.013$). Indeed, female reviewers from lower-ranked institutions report *lower* confidence

[22]Situations that may invite greater scrutiny by organizers might also increase women's self-doubt (Sarsons and Xu, 2021; Bordalo et al., 2019), leading them to exert greater effort and, in turn, report greater confidence, independently of organizers' behavior. This alternative mechanism is, however, unlikely, as we argue below that effort-based explanations are not supported by the data.

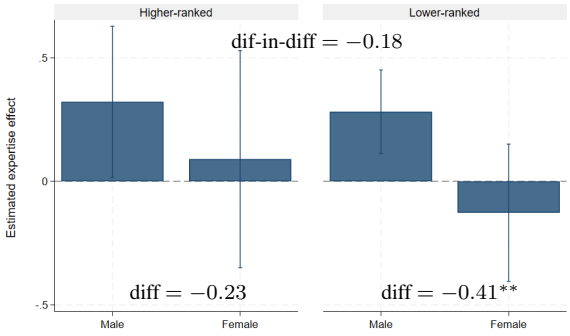
[23]The reviewer's PhD institution is classified as "lower-ranked" if it is not among the following: Harvard, MIT, Yale, Northwestern, Columbia, Penn, LSE, UCL, Oxford, or Warwick. Reviewers from these top institutions account for around 13% of all reviews in our dataset (15% for male reviewers and 9% for female reviewers).

[24]In both analyses we exclude reviewers who are organizers themselves. Since the organizing committee changes each year, this comparison does not involve the same set of individuals across conferences.

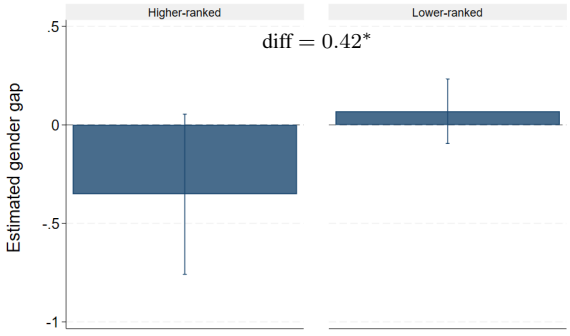
when reviewing within rather than outside their field of expertise, although the point estimate is not statistically significant.

These gender differences in how institutional rank relates to confidence translate into a *more positive* gender confidence gap among reviewers from lower-ranked institutions. Female reviewers report, on average, 0.07 points ($p = 0.41$) higher confidence than male reviewers, compared with a negative and significant gender gap among reviewers from higher-ranked institutions (-0.35, $p = 0.09$), as shown in panel b.[25] In line with the patterns described above, Appendix Table A.12 shows that the gender gap turns significantly positive for reviewers from lower-rank institutions reviewing papers outside their field of expertise (0.23, $p = 0.03$). On the contrary, the traditional negative gender gap (male reviewers stating higher confidence) appears when reviewers from higher-ranked institutions review papers within their field of expertise, where there are likely lower expectations of credibility questioning and less uncertainty (-0.49, $p = 0.04$).

Figure 7- The role of lower-ranked PhD institutions



(a) Estimated expertise effect by PhD institution rank and reviwer's gender



(b) Estimated gender gap by PhD institution rank

Notes: Bar graph that displays the estimated expertise effect on confidence (within minus outside the field of expertise) by reviewer's PhD institution rank and gender (panel a) and the estimated gender confidence gap (female minus male) by reviewer's PhD institution rank (panel b), net of paper and reviewer controls as specified in equation (1) and excluding reviewers who are organizers. A reviewer is considered to have a PhD from a higher-ranked institution if they got their PhD from Harvard, MIT, Yale, Northwestern, Columbia, Penn, LSE, UCL, Oxford, or Warwick. The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full regressions are shown in Table A.12.

Familiarity with the organizers Familiarity with the organizers is also strongly associated with confidence, but in opposite directions for male and female reviewers. Female reviewers report, on average, 0.31 points ($p = 0.014$) higher confidence when they are not colleagues than when they are. For male reviewers, the relationship is significantly weaker and reversed: non-colleagues

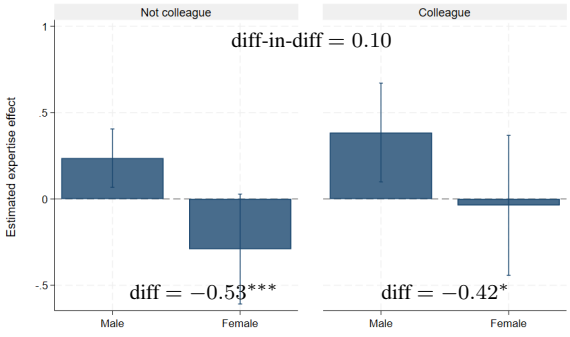
[25]This result is distinct from the result on rank, as reviewers from lower-ranked institutions are not more likely to be Professors. If anything, the contrary is true.

report 0.07 points ($p = 0.087$) lower confidence (Appendix Table A.13).

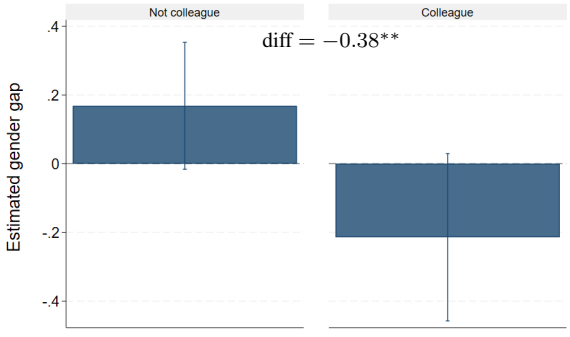
Importantly, familiarity with the organizers affects how closely confidence reflects expertise for women, but not for men. Figure 8, panel a shows that among male reviewers, reviewing papers within their field of expertise is associated with higher confidence regardless of whether they are familiar with the organizers. Women, on the other hand, report significantly lower confidence when reviewing papers within the field of expertise when they are not familiar with the organizers (-0.29, $p = 0.07$), while topic-specific expertise has a non-significant effect when they know the organizers (-0.04, $p = 0.86$). This results in a larger gender difference in the expertise effect among non-colleagues (-0.53, $p = 0.004$) than among colleagues (-0.42, $p = 0.095$).

Overall, these patterns translate into a positive gender confidence gap among non-colleagues. Female reviewers report, on average, 0.17 points ($p = 0.074$) higher confidence than male reviewers, compared with a negative gender gap among colleagues (-0.21, $p = 0.085$), as shown in panel b (difference $= -0.38$, $p = 0.013$). In line with the heterogeneity discussed in the previous paragraph, Appendix Table A.13 additionally shows that the positive gender difference in confidence is driven by non-colleagues who review papers outside their field of expertise (0.40, $p = 0.001$), i.e. exactly when the organizers have less direct information about reviewers' expertise, and women might expect their expertise to be questioned. When colleagues review papers within their field of expertise—where expertise is more readily observable, and uncertainty is lower—the negative gender difference, with women reporting lower confidence than men, reappears (-0.47, $p = 0.020$).

Figure 8- The role of familiarity to organizers



(a) Estimated expertise effect by colleague status and reviewer's gender



(b) Estimated gender gap by colleague status

Notes: Bar graph that displays the estimated expertise effect on confidence (within minus outside the field of expertise) by reviewer's familiarity with the organizers and gender (panel a) and the estimated gender confidence gap (female minus male) by reviewer's familiarity with the organizers (panel b), net of paper and reviewer controls as specified in equation (1) and excluding reviewers who are organizers. Reviewers familiar with the organizers are those working at the same institution as the organizers (colleagues). The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full regressions are shown in Table A.13.

These reviewer-level patterns suggest that women's confidence responds differently to situations in which their expertise is less readily observable and may be subject to greater scrutiny. Female reviewers report relatively higher confidence when they come from lower-ranked institutions or are less familiar to organizers, and their confidence less closely reflects topic-specific expertise in these settings. The fact that the positive gender confidence gap is concentrated in contexts where women's expertise may be more closely scrutinized is consistent with anticipated scrutiny playing a role in women's higher stated confidence.

VC Reviewing Effort

As a last exercise, we turn to reviewers' commenting behaviour to assess the role of effort. This matters because expectations of heightened scrutiny could affect confidence through different channels such as greater effort, or incentives to signal competence by reporting higher confidence without necessarily increasing effort. We examine this distinction using reviewers' commenting behaviour. Reviewers are asked to leave a comment for the organizers before submitting. Under the assumption that more diligent reviewers are more likely to leave feedback, we can use commenting behavior as a proxy for effort. We use reviewers' comments in two ways.

First, we examine the probability of leaving a comment. Overall, 54.06% of reviewers leave a written comment, with men more likely to do so—a difference robust to controlling for paper and reviewer characteristics (female coefficient -0.13; $p = 0.000$, Appendix Table A.14). Commenting behavior follows the opposite pattern of confidence: men are more likely to leave comments when reviewing papers outside their field, whereas no significant gender gap is present for papers within the field of expertise (-0.18, $p = 0.000$ vs -0.05, $p = 0.38$, Appendix Table A.14), a pattern observed across all ranks.

We additionally ask Generative AI to classify comments based on the level of effort put into the review on a scale from 1-5.[26] Given the reviewer leaves a comment, the resulting effort scores are strongly correlated with comment length ($\rho = 0.731, p < 0.000$), providing evidence that the scores capture variation in the extent of reviewing effort. We observe the same patterns as above using this measure (Appendix Table A.14).

The evidence on commenting behaviour suggests that women's higher confidence in the contexts identified above is not accompanied by greater reviewing effort, making a simple effort-based explanation difficult to reconcile with the observed gender difference in confidence. Instead, our results suggest that stated confidence reflects factors other than self-assessment of expertise. Whether this reflects signaling of expertise or another mechanism remains an open question. Nonetheless,

[26]ChatGPT prompt: "Can you go through the list of comments from reviewers of conference submissions to conference organizers. Analyse and give a score for level of effort put into the review (1-5)?". No comment is classified as "0: no effort".

the patterns are consistent with women anticipating greater scrutiny of their competence and responding accordingly.

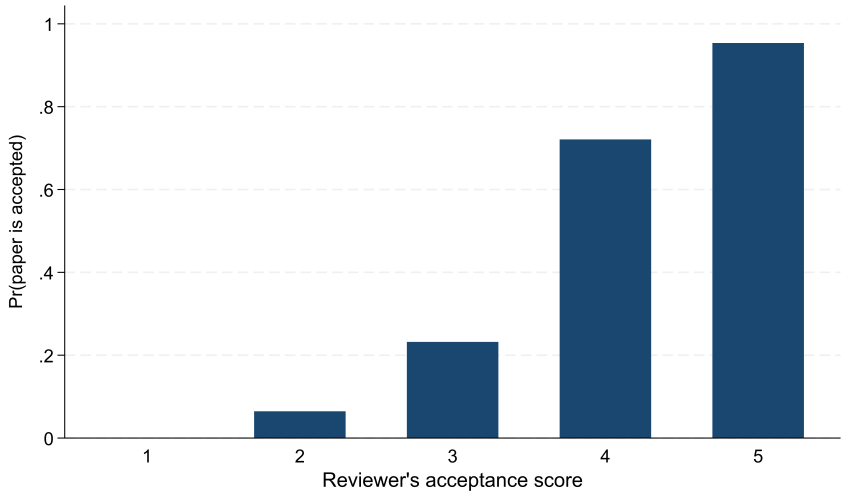
VI The Gender Influence Gap

The above result raises a natural question: are reviewers' assessments in fact weighted differently in the final decision? To examine this, we study how organizers map reviewers' acceptance scores into acceptance decisions, and whether this mapping differs by reviewer gender.

The acceptance score assigned by reviewers is a strong predictor of organizers' decisions, as shown in Figure 9. Papers with a score of 2 or lower ("do not accept" or "weak reject") are rarely accepted, with acceptance rates of 0% and 6%, respectively. At the other end, papers receiving a score of 5 are almost always accepted (95%), and those with a score of 4 have a high acceptance rate (72%). By contrast, papers with a score of 3 ("borderline accept/reject") are accepted only 23% of the time.

This distribution of acceptance decisions reveals a clear pattern in how much discretion organizers exercise. Organizers have relatively little discretion when papers receive very strong or very weak recommendations, as acceptance decisions closely track reviewer scores. By contrast, a score of 3 provides a more ambiguous signal of paper quality, leaving greater scope for organizer judgment. This is therefore the setting in which differences in how reviewers' recommendations are weighted are most likely to emerge.

Figure 9- Probability of the paper being accepted by the reviewer's acceptance score



Notes: Share of papers accepted to the conference (y axis) by reviewers' acceptance scores (x-axis).

We next examine whether this mapping differs by reviewer gender. Specifically, we compare

acceptance decisions for papers receiving the same reviewer score, controlling for an array of paper and reviewer characteristics:

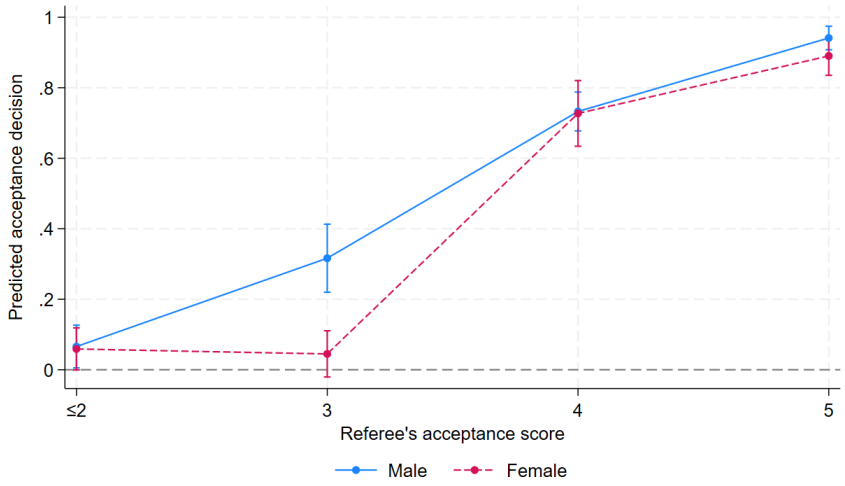
$$Y_{ipt} = \alpha_t + \sum_{j=2}^{j=5} \beta_{1,j}[\mathbb{1}(score = j)] + \sum_{j=2}^{j=5} \beta_{2,j}[\mathbb{1}(score = j)] \times Female_i + X_i'\gamma + X_p'\delta + \epsilon_{ipt} \quad (2)$$

where Y_{ipt} denotes a binary indicator that organizers in year t accept paper p for which reviewer i provided acceptance score j. $\mathbb{1}(score = j)$ is a dummy equal to one for each score value j. Papers with scores equal to one and two are aggregated ($j = 2$) to ensure a meaningful group size. We include year FEs (α_t) and control for the same set of reviewer-specific and paper-specific characteristics as described for equation (1). In the figures that follow, we plot the predicted probability of acceptance by reviewer score and gender, net of these controls.

The evidence suggests that organizers place different weights on reviewers' recommendations depending on reviewers' gender. Four observations support this interpretation.

First, on average, organizers' decisions align less closely with female reviewers' recommendations than with male reviewers' for marginal submissions—the key margin where reviewer evaluations are most informative for the final decision. Figure 10 shows that for papers receiving a score of 3 ("borderline accept/reject"), the probability of acceptance is higher when the reviewer is male. By contrast, we find no meaningful differences for papers receiving more decisive recommendations (scores 1–2 or 4–5), where acceptance decisions are largely mechanical. A formal test of these differences is reported in Table A.15.

Figure 10- Gender gap in the probability of the paper being accepted

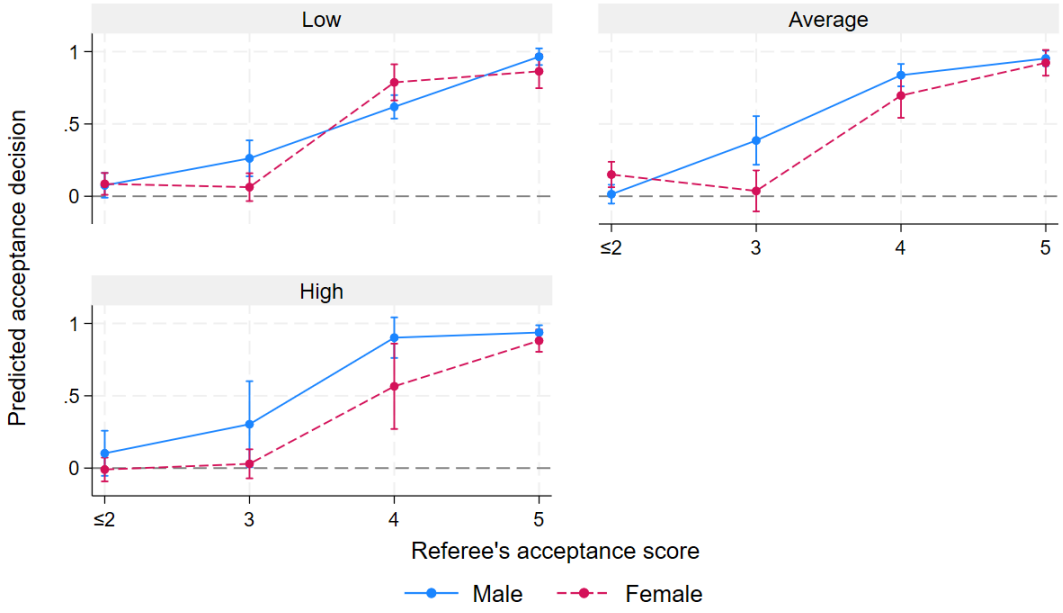


Notes: Average predicted probability that the paper is accepted by the organizers, plotted against reviewer's acceptance decisions, net of reviewer and paper characteristics as in equation (2). The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full results are reported in Table A.15; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

If organizers perfectly followed the reviewers' decisions, they would accept papers with 50% probability, a random draw, when reviewers are undecided (score = 3). We assume that reviewers declaring indecision are genuinely uncertain about whether the paper should be accepted to the conference, equivalent to a coin flip if forced to choose. Under this assumption, the 50% acceptance rate can be used as the natural benchmark that would arise if organizers followed reviewers' evaluations on average. Deviations from this benchmark therefore indicate the extent to which organizers' decisions diverge from the recommendations of reviewers who report being undecided. The observed differences—18 percentage points for male reviewers and 46 percentage points for female reviewers—suggest greater divergence from female reviewers' evaluations.

Second, the gap is larger among reviewers who report higher confidence, as shown in Figure 11 (estimates in Table A.16). In particular, a negative gender gap also emerges for papers receiving a score of 4 ("probably accept") among reviewers with average and high confidence. This suggests that the differential weighting persists even when female reviewers report high confidence in their assessments.

Figure 11- Gender gap in the probability of the paper being accepted - by confidence



Notes: Average predicted probability that the paper is accepted by the organizers, plotted against reviewers' acceptance decisions, net of reviewer and paper characteristics as in equation (2). Each graph displays results for a subsample of reviewers with different confidence: reviewers whose confidence score is less than or equal to 3 (top left), equal to 4 (top right), and equal to 5 (bottom left), respectively. The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full results are reported in Table A.16; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

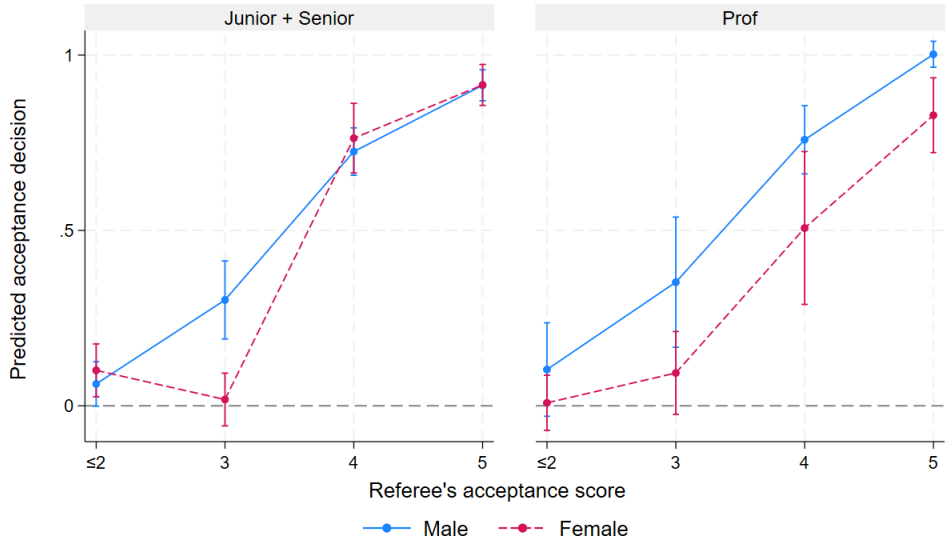
Third, these results are unlikely to be explained by systematic differences in scoring behavior. Table A.19 shows no evidence that female reviewers assign higher acceptance scores than male reviewers. An alternative explanation is that organizers *believe* female reviewers to be more lenient and therefore discount their recommendations. However, for this mechanism to account for the patterns seen in Figure 11, organizers would need to perceive female reviewers as increasingly lenient as their expressed confidence increases. We have no particular reason to expect such beliefs.

Lastly, the differential weighting of reviewers' recommendations is particularly pronounced among Professors. As shown in Figure 12, recommendations by female professors are less closely aligned with organizers' decisions than those by male professors for papers receiving acceptance scores of 3 or above.[27] This pattern contrasts with prior work suggesting that stronger credentials

[27]The gender gap becomes less precisely estimated when we restrict the sample to Professors, as the number of observations falls to 254. Also in this case, however, the results are unlikely to be explained by female professors being systematically more lenient. Table A.19 shows no evidence that female professors assign higher acceptance scores than male professors. If anything, consistent with the higher stated confidence, female professors appear to be more decisive, as they are less likely to state they are uncertain about accepting or rejecting a paper (acceptance score=3).

may mitigate gender differences in how expertise is evaluated.[28] Here, differential weighting is

Figure 12- Gender gap in the probability of the paper being accepted - by rank



Notes: Average predicted probability that the paper will be accepted by the organizers by reviewer's acceptance decisions and gender, net of reviewer and paper characteristics as in equation (2). Each graph displays results for a subsample of reviewers: up to and including Associate Professors or equivalent (left) and Professors or equivalent (right). The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals using robust standard errors. Full results are reported in Table A.16; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

most pronounced among the most senior reviewers. This may help explain why the gender confidence gap is also largest among professors (Figure 2): after longer careers in the profession, female professors may be more aware of the greater scrutiny they face and respond by reporting higher confidence.

These results indicate that in contexts where there is greater scope for discretion—such as for borderline submissions—female reviewers' expertise appears to be weighted less. This may, in turn, help explain why female reviewers behave differently in the settings examined in Section V, reporting higher confidence when uncertainty about their expertise is more likely to arise. If female reviewers anticipate that their expertise will receive greater scrutiny or carry less weight in such contexts, higher reported confidence may represent a response to these incentives rather than simply a different assessment of their own expertise.

[28] If women face discrimination, successful women may be perceived as exceptionally capable (Bohren et al., 2019). Consistent with this, the gender gap in teaching evaluations is smaller among more senior professors (Mengel et al., 2019), and expert opinions can be more persuasive when expressed by senior female economists (Sievertsen and Smith, 2025a), potentially reflecting the salience of their credentials.

VIA Implications for Paper Outcomes

The evidence above shows that, despite reporting significantly higher confidence, female reviewers' evaluations appear to be weighted differently by organizers, with decisions deviating more from their recommendations.

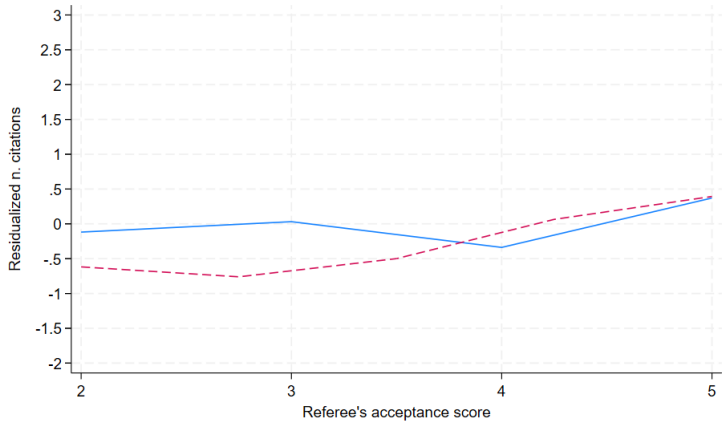
This raises the question of whether such differential weighting affects the quality of final decisions. To examine this, we study the relationship between reviewers' acceptance scores and subsequent paper outcomes, proxied by citations. We consider three measures: (i) citations accumulated in the year following the conference, (ii) total citations by January 2025, and (iii) total citations by January 2025 for papers submitted before 2022, for which longer-run outcomes can be observed.

Female reviewers' evaluations are more strongly associated with subsequent paper success. Figure 13 shows that the relationship between acceptance scores and citations is steeper for female reviewers than for male reviewers. Among low-scoring papers, those evaluated by female reviewers receive fewer citations, while among high-scoring papers, they receive more. These differences are small in the short run but become more pronounced over time.

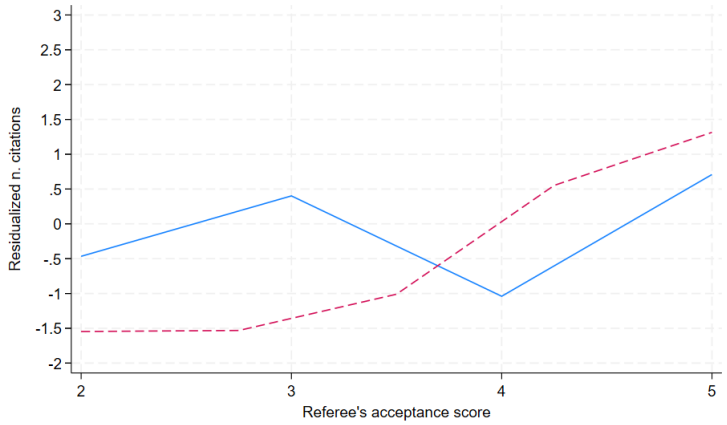
These patterns are unlikely to be driven solely by differences in acceptance rates between male- and female-reviewed papers. Differences in citations for low-scoring papers, particularly those with a score of 3, may partly reflect selection into acceptance, as papers reviewed by male reviewers are more likely to be accepted and therefore receive greater exposure. However, the divergence in citation patterns widens over time, when the direct influence of conference acceptance is likely to diminish and citations may more closely reflect underlying paper quality. A similar pattern emerges among high-scoring papers, where papers evaluated by female reviewers accumulate more citations over time.

Our findings indicate that female reviewers' assessments are at least as informative—and in some cases more predictive—of subsequent paper outcomes as those of male reviewers. This suggests that their expertise may be undervalued in the decision-making process.

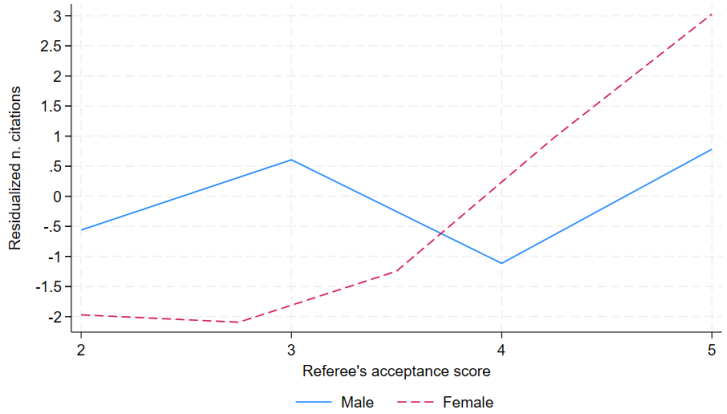
Figure 13- Citations post conference



(a) Year after the conference



(b) By January 2025



(c) By January 2025 (Including conference years up to 2021)

Notes: This figure plots the residuals from a regression of the paper citations post-conference on paper and reviewer characteristics (as specified in equation (2)) against reviewer's acceptance scores. The lines represents the smoothed values of a kernel local polynomial regression.

VII Discussion and Concluding Remarks

In this paper, we document gender differences in confidence and influence in a novel and consequential setting: the peer review of academic conference submissions. Using data from the Irish Economic Association (IEA) Conference from 2017–2023, we show that women's stated confidence responds differently to contextual cues about expertise than men's. In particular, women report relatively higher confidence in settings characterized by greater uncertainty about how their expertise may be perceived or evaluated.

While the magnitude of the effects varies across settings, the direction of the gender difference in the response to these contextual cues is remarkably consistent. This consistency is difficult to reconcile with the view that female reviewers in our sample are simply an unusually confident group of women who differ fundamentally from those studied in previous research. Indeed, the same women exhibit confidence patterns closer to conventional findings in the literature when the conditions associated with greater uncertainty are absent. The evidence therefore points to context, rather than a fixed difference in confidence, as an important determinant of how women report their confidence.

Higher stated confidence in the presence of heightened uncertainty is consistent with a response to anticipated scrutiny. Our evidence provides suggestive support for this interpretation: the gender difference in confidence is larger when women's expertise is less readily observable to organizers. We then consider whether anticipated scrutiny could increase confidence through greater reviewing effort. We find little evidence that women's higher confidence is accompanied by greater effort, which weighs against this channel. Rather, our findings suggest that stated confidence reflects considerations beyond a genuine self-assessment of expertise. Whether this stems from signaling or some other mechanism is a question we leave open.

We further show that female reviewers' concerns about how their expertise is evaluated may be well-founded: their recommendations are weighted differently in the final decision. Whether differences in stated confidence contribute to this influence gap or instead reflect a response to it remains unclear. Organizers may partially respond to females' higher confidence, which could reduce the influence gap (or explain why it is not larger); ignore it; or place less weight on evaluations if higher confidence is anticipated (see e.g. Demiral and Mollerstrom, 2024). What we can show is that female reviewers' evaluations are at least as informative, and in some cases more predictive, of subsequent citations than those of male reviewers. Their assessments therefore appear to contain at least as much information about paper quality as those of male reviewers, despite being weighted differently in the final decision.

Although our paper focuses on the Irish Economics Association Conference, there is no reason to expect reputational concerns and expectations of scrutiny to be unique to this setting. Women

make up roughly one-third of economists in Ireland, a share broadly in line with the rest of Europe.[29] Moreover, the Irish economics community and its flagship conference are relatively small,[30] which could shape reputational dynamics in opposing ways. On one hand, a smaller community raises the reputational stakes for reviewers, who are more likely to know organizers personally or interact with them again in the future, making concerns about how one's expertise is perceived especially salient. On the other hand, in larger, more anonymous settings—where evaluators have less information about one another's expertise—reliance on priors and stereotypes may be more pronounced, similarly heightening concerns about scrutiny in expertise evaluation. Finally, acceptance or rejection may carry higher career stakes for authors at larger, more prestigious conferences, potentially increasing the stakes of reviewers' evaluations and the incentives surrounding the expression of confidence.

Finally, although our analysis focuses on economists and conference acceptance decisions, the underlying features of this environment are common to many evaluation settings. When evaluators' expertise is imperfectly observed and decision-makers retain discretion, both the expression of confidence and the weight placed on evaluations may differ systematically across individuals. These considerations are relevant in a wide range of domains, including hiring and promotion decisions, the assessment of loan applicants, and the provision of tenant references, where evaluators' expertise is not fully known and their assessments are interpreted alongside signals such as confidence.

References

Abraham, D., B. Greiner, and M. Stephanides (2023). On the internet you can be anyone: An experiment on strategic avatar choice in online marketplaces. *Journal of Economic Behavior & Organization 206*, 251–261.

Adams, R. and P. Funk (2012). Beyond the glass ceiling: Does gender matter? *Management Science 58*(2), 219–235.

Aksoy, B., I. Chadd, and B. H. Koh (2023). Sexual identity, gender, and anticipated discrimination in prosocial behavior. *European Economic Review 154*, 104427.

Angeli, D., I. Matavelli, and F. Secco (2026). Expected discrimination and job search. Working paper.

Arai, M. and P. Skogman Thoursie (2009). Renouncing personal names: An empirical examination of surname change and earnings. *Journal of Labor Economics 27*(1), 127–147.

Ash, E., D. L. Chen, and A. Ornaghi (2024, January). Gender attitudes in the judiciary: Evidence from us circuit courts. *American Economic Journal: Applied Economics 16*(1), 314–50.

Ashenfelter, O. and e. Albert Rees (1973). *Discrimination in Labor Markets*. Princeton University Press.

[29] Women represent approximately 33% of economists in Ireland, broadly in line with France and Italy (32% each) and the European average (32.5%) reported in Auriol et al. (2022), though somewhat higher than in the UK (26%) or the US (23%) according to IDEAS.

[30] The number of economists in Ireland is approximately 345, compared to 3,848 in France, 3,517 in Italy, 3,954 in the UK, and 12,611 in the US, based on IDEAS.

Auriol, E., G. Friebel, A. Weinberger, and S. Wilhelm (2022). Underrepresentation of women in the economics profession more pronounced in the United States compared to heterogeneous Europe. *Proceedings of the National Academy of Sciences 119*(16), e2118853119.

Baltrunaite, A., A. Casarico, and L. Rizzica (2025). Women in economics: The role of gendered references at entry in the profession. *European Economic Review 180*, 105155.

Bandiera, O., N. Parekh, B. Petrongolo, and M. Rao (2022). Men are from mars, and women too: A Bayesian meta-analysis of overconfidence experiments. *Economica 89*, S38–S70.

Barber, B. M. and T. Odean (2001, 02). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics 116*(1), 261–292.

Batistich, M. K., T. N. Bond, S. Linde, and K. J. Mumford (2024, Aug). Statistical discrimination and optimal mismatch in college major selection. IZA Discussion Paper 17237, Institute of Labor Economics (IZA), Bonn.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology 59*(5), 960.

Biavaschi, C., C. Giulietti, and Z. Siddique (2017). The economic payoff of name Americanization. *Journal of Labor Economics 35*(4), 1089–1116.

Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review 109*(10), 3395–3436.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). Beliefs about gender. *American Economic Review 109*(3), 739–773.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics 145*, 27–41.

Boring, A. and A. Philippe (2021). Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching. *Journal of Public Economics 193*, 104323.

Bostwick, V. K. and B. A. Weinberg (2022). Nevertheless she persisted? Gender peer effects in doctoral stem programs. *Journal of Labor Economics 40*(2), 397–436.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *Quarterly Journal of Economics 129*(3), 1409–1447.

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2020). Are referees and editors in economics gender neutral? *Quarterly Journal of Economics 135*(1), 269–327.

Charness, G., R. Cobo-Reyes, S. Meraglia, and Á. Sánchez (2020). Anticipated discrimination, choices, and performance: Experimental evidence. *European Economic Review 127*, 103473.

Coffman, K., M. R. Collis, and L. Kulkarni (2024). Stereotypes and belief updating. *Journal of the European Economic Association 22*(3), 1011–1054.

Coffman, K. B. (2014, 09). Evidence on self-stereotyping and the contribution of ideas. *Quarterly Journal of Economics 129*(4), 1625–1660.

Coffman, K. B., C. L. Exley, and M. Niederle (2021). The role of beliefs in driving gender discrimination. *Management Science 67*(6), 3551–3569.

de Vaan, M. and T. Stuart (2022). Gender in the markets for expertise. *American Sociological Review 87*(3), 443–477.

Demiral, E. E. and J. Mollerstrom (2024). Signaling confidence. *Journal of Economic Behavior & Organization 226*, 106691.

Doleac, J. L., E. Hengel, and E. Pancotti (2021). Diversity in economics seminars: Who gives invited talks? *AEA Papers and Proceedings 111*, 55–59.

Dupas, P., A. Handlan, A. S. Modestino, M. Niederle, M. Seré, H. Sheng, J. Wolfers, S. D. Collective, et al. (2025). Gender differences in economics seminars. *American Economic Review*.

Eberhardt, M., G. Facchini, and V. Rueda (2023). Gender differences in reference letters: Evidence from the economics job market. *The Economic Journal 133*(655), 2676–2708.

Exley, C. L., R. Fisman, J. B. Kessler, L.-P. Lepage, X. Li, C. Low, X. Shan, M. Toma, and B. Zafar (2024). Information-optional policies and the gender concealment gap. National Bureau of Economic Research Working Paper No. 32350.

Exley, C. L. and J. B. Kessler (2022). The gender gap in self-promotion. *Quarterly Journal of Economics 137*(3), 1345–1381.

Exley, C. L. and K. Nielsen (2024, March). The gender gap in confidence: Expected but not accounted for. *American Economic Review 114*(3), 851–85.

Gordon, R. and G. B. Dahl (2013). Views among economists: Professional consensus or point-counterpoint? *American Economic Review 103*(3), 629–635.

He, H., D. Neumark, and Q. Weng (2024). Gender reveals in the labor market: Evidence on gender signaling and statistical discrimination in an online health care market. Working Paper 32929, National Bureau of Economic Research.

Hebert, C. (2025). Gender stereotypes and entrepreneur financing. *The Review of Financial Studies*, hhaf049.

Hengel, E. (2022). Publishing while female. *The Economic Journal 132*(648), 2951–2991.

Hengel, E. and E. Moon (2020). Gender and equality at top economics journals. Working Paper.

Hospido, L. and C. Sanz (2021). Gender gaps in the evaluation of research: Evidence from submissions to economics conferences. *Oxford Bulletin of Economics and Statistics 83*(3), 590–618.

Jakobsson, N. (2012, July). Gender and confidence: Are women underconfident? *Applied Economics Letters 19*(11), 1057–1059.

Kang, S. K., K. A. DeCelles, A. Tilcsik, and S. Jun (2016). Whitened résumés: Race and self-presentation in the labor market. *Administrative Science Quarterly 61*(3), 469–502.

Klein, G., Z. Shtudiner, and M. Zwilling (2021). Uncovering gender bias in attitudes towards financial advisors. *Journal of Economic Behavior & Organization 189*, 257–273.

Koffi, M. (2021). Gendered citations at top economic journals. *AEA Papers and Proceedings 111*, 60–64.

Kudashvili, N. and P. Lergetporer (2022). Minorities' strategic response to discrimination: Experimental evidence. *Journal of Public Economics 208*, 104630.

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2021). Choice of majors: Are women really different from men? *Economics of Education Review 81*, 102079.

Lane, T. (2024). The strategic use of social identity. *Journal of Economic Behavior & Organization 224*, 355–368.

Lang, K. and M. Manove (2011, June). Education and labor market discrimination. *American Economic Review 101*(4), 1467–96.

Lee, T., M. Morelli, M. Pappalettera, D. Sansone, and S. Sardoschau (2025). EEA professional climate survey report. Available at https://doi.org/10.48550/arXiv.2508.04302.

Lepage, L.-P., X. Li, and B. Zafar (2025, April). Anticipated discrimination and major choice. Working Paper 33680, National Bureau of Economic Research.

Lundberg, S. and J. Stearns (2019). Women in economics: Stalled progress. *Journal of Economic Perspectives 33*(1), 3–22.

Lundeberg, M. A., P. W. Fox, and J. Punćcohaŕ (1994). Highly confident, but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology 86*, 114–121.

Mazei, J., J. Hüffmeier, P. A. Freund, A. F. Stuhlmacher, L. Bilke, and G. Hertel (2015). A meta-analysis on gender differences in negotiation outcomes and their moderators. *Psychological Bulletin 141*(1), 85.

Mengel, F., J. Sauermann, and U. Zölitz (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association 17*(2), 535–566.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2022). Managing self-confidence: Theory and experimental evidence. *Management Science 68*(11), 7793–7817.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review 62*(4), 659–661.

Radbruch, J. and A. Schiprowski (2023). Committee deliberation and gender differences in influences. CRC Discussion Paper 398.

Reubeck, H. (2025). Causes and consequences of perceived workplace discrimination. Working paper.

Reuben, E., P. Sapienza, and L. Zingales (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences 111*(12), 4403–4408.

Rudman, L. A. and J. E. Phelan (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior 28*, 61–79.

Sarsons, H. (2017). Interpreting signals in the labor market: Evidence from medical referrals. Working Paper.

Sarsons, H., K. Gërxhani, E. Reuben, and A. Schram (2021). Gender differences in recognition for group work. *Journal of Political Economy 129*(1), 101–147.

Sarsons, H. and G. Xu (2021). Confidence men? Evidence on confidence and gender among top economists. *AEA Papers and Proceedings 111*, 65–68.

Sievertsen, H. H. and S. Smith (2025a). Do female experts face an authority gap? Evidence from economics. *Journal of Economic Behavior & Organization 231*, 106912.

Sievertsen, H. H. and S. Smith (2025b). The gender gap in expert voices: Evidence from economics. *Public Understanding of Science 34*(4), 446–458.

Ulgade, P. (2022). Gender, grade sensitivity, and major choice. Working paper.

Zanella, M. (2024). Stereotypical Selection. Trinity Economics Papers 0224.

Zussman, A. (2013). Ethnic discrimination: Lessons from the israeli online market for used cars. *The Economic Journal 123*(572), F433–F468.

Online Appendix

A Additional Tables and Figures

Table A.1- Origin of Submitters

	Submitters	
	N.	Percent
Central Bank of Ireland	106	11.78
Trinity College Dublin	88	9.78
University College Dublin	87	9.67
The Economic and Social Research Institute (ESRI)	85	9.44
University College Cork	43	4.78
University of Galway	32	3.56
Queen's University Belfast	22	2.44
Technological University Dublin	21	2.33
Maynooth University	19	2.11
University of Limerick	13	1.44
European Central Bank	10	1.11
Irish Fiscal Advisory Council (IFAC)	10	1.11
Other within Ireland (<10)	29	3.22
Other outside Ireland (<10)	335	37.22
Total	**900**	**100**

Notes: Descriptive statistics on the full sample of submissions to the IEA Conference 2017-2023.

Table A.2- Reviewers' Characteristics (I)

	Reviewers	
	N.	Percent
Panel A: Current Institution		
University College Dublin	55	17.57
Central Bank of Ireland	52	16.61
The Economic and Social Research Institute	48	15.34
University College Cork	40	12.78
University of Galway	39	12.46
Maynooth University	22	7.03
University of Limerick	20	6.39
Trinity College Dublin	18	5.75
Dublin City University	8	2.56
European Central Bank	2	0.64
Irish Fiscal Advisory Council	2	0.64
Manchester University	2	0.64
International Monetary Fund	1	0.32
European Stability Mechanism	1	0.32
The London School of Economics	1	0.32
OECD	1	0.32
Queen's University Belfast	1	0.32
Panel B: PhD Institution		
Harvard, MIT, Yale, Northwestern, Columbia, Penn	29	9.27
LSE, UCL, Oxford, Warwick	18	5.75
TCD, UCD	88	28.12
Other Ireland	58	18.53
Other Europe	71	22.68
Other US	29	9.27
Other	9	2.88
Missing	11	3.51
Panel C: Rank		
1: Ph.D. student or equivalent	8	2.56
2: Post-doc, Research Fellow or equivalent	17	5.43
3: Assistant Professor or equivalent	108	34.50
4: Associate Professor or equivalent	87	27.80
5: Professor or equivalent	87	27.80
Missing	6	1.92
Total	**313**	**100**

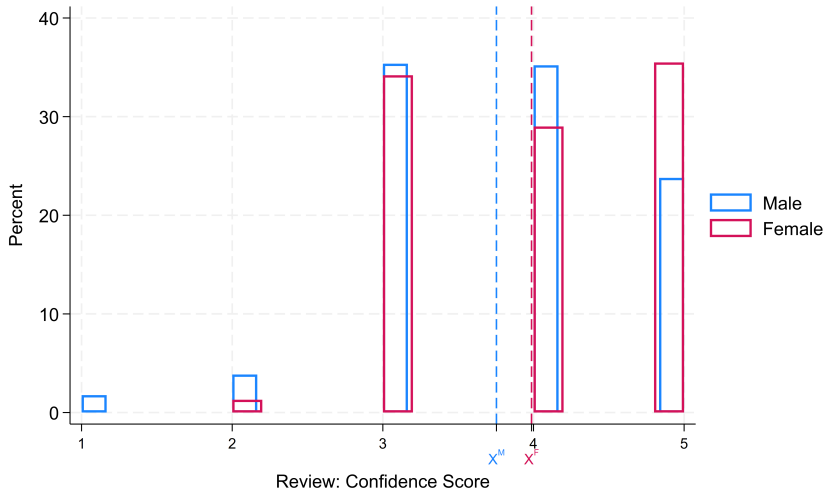
Notes: Descriptive statistics on reviewer-year pairs for the IEA Conference 2017–2023. For non-academic reviewers, ranks are assigned as follows: rank 1 (Economist), rank 2 (Senior Economist, Research Officer), rank 3 (Principal Economist, Manager, Advisor, Senior Advisor, Senior Research Officer), rank 4 (Head of Function, Head of Division), and rank 5 (Director, Deputy Governor). This conversion was provided by the ESRI and the Central Bank of Ireland, the primary affiliations for non-academic reviewers.

Table A.3- Reviewers' Characteristics (II)

	N.	Mean	Std. Dev.	Min	Max
Female	313	0.28	0.45	0	1
Years post-Phd	313	15.59	8.27	-1*	49
N. Reviews (current year)	313	2.88	3.58	1	36
N. Reviews (excluding current)	313	5.71	6.57	0	39
N. Citations	313	1120.25	1138.42	0	7050
N. Solo-authored publications	313	4.02	9.93	0	152
N. Co-authored publications	313	17.48	15.93	0	119

Notes: Descriptive statistics on reviewer-year pairs for the IEA Conference 2017–2023. *One of the reviewers was a PhD student at the time of reviewing for the conference. The number of publications, citations, and years post-PhD are imputed to average if we could not find this information about the reviewer (31, 1, 30 reviewer-year pairs, respectively). The results are robust to replicating the analysis excluding these observations.

Figure A.1- Reviewers' confidence scores by gender



Notes: Histogram of the reviewers' confidence scores by gender. The vertical lines indicate the average confidence scores for the two groups: X^M for male and X^F for female.

Table A.4 - Gender gap in declared confidence: full model

	Reviewers' confidence score {1;5}				
	(1)	(2)	(3)	(4)	(5)
Female	0.19***	0.22***	0.19**	0.18**	0.21**
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
Economic Development and Growth		0.19	0.13	0.11	0.13
		(0.30)	(0.28)	(0.28)	(0.30)
Economic History		1.01***	0.92***	0.97***	1.22***
		(0.33)	(0.30)	(0.30)	(0.30)
Energy Economics		0.05	-0.03	-0.06	0.02
		(0.41)	(0.39)	(0.36)	(0.37)
Financial Economics		0.26	0.30	0.26	0.39
		(0.26)	(0.25)	(0.25)	(0.26)
Health, Education, and Welfare Economics		0.34	0.28	0.32	0.40
		(0.25)	(0.25)	(0.24)	(0.25)
History of Economic Thought		-0.36	-0.12	-0.08	-0.23
		(0.25)	(0.26)	(0.27)	(0.26)
Housing Economics		0.51	0.66**	0.41	0.37
		(0.32)	(0.26)	(0.26)	(0.29)
Industrial Organisation		0.32	0.31	0.32	0.33
		(0.31)	(0.30)	(0.30)	(0.31)
International Economics		0.51*	0.41	0.41*	0.43*
		(0.26)	(0.25)	(0.24)	(0.26)
Labour/Demographic Economics		0.53**	0.47*	0.50**	0.63**
		(0.26)	(0.25)	(0.24)	(0.26)
Macroeconomics		0.42*	0.40*	0.34	0.33
		(0.25)	(0.24)	(0.23)	(0.24)
Microeconomics		0.20	0.29	0.29	0.27
		(0.25)	(0.24)	(0.23)	(0.24)
Public Economics		0.08	0.22	0.21	0.24
		(0.26)	(0.25)	(0.24)	(0.25)
Regional/Real Estate/Transport Economics		0.38	0.43	0.36	0.43
		(0.31)	(0.27)	(0.26)	(0.29)
Teaching Economics		0.21	0.31	0.39	1.24***
		(0.39)	(0.38)	(0.40)	(0.30)
Blinded		-0.02	-0.03	-0.02	-0.17
		(0.12)	(0.13)	(0.14)	(0.16)
Prop. authors male		-0.04	-0.06	-0.05	-0.06
		(0.08)	(0.08)	(0.08)	(0.09)
Prop. authors PhD		0.13	0.15*	0.14*	0.14
		(0.09)	(0.08)	(0.08)	(0.09)
Prop. authors Irish		0.07	0.09	0.10	0.09
		(0.10)	(0.10)	(0.10)	(0.12)

Continued on next page

Table A.4 – continued from previous page

	(1)	(2)	(3)	(4)	(5)
Prop. authors Irish WP		0.20*	0.17*	0.16	0.19
		(0.10)	(0.10)	(0.10)	(0.12)
Number of authors		0.05*	0.04	0.04	0.05
		(0.03)	(0.03)	(0.03)	(0.03)
Same institution DV		0.20	0.16	0.13	0.16
		(0.14)	(0.13)	(0.13)	(0.14)
Paper pre-citations		0.02	0.02	0.02	0.02
		(0.01)	(0.01)	(0.01)	(0.01)
No paper citation DV		0.15*	0.19**	0.18**	0.19**
		(0.09)	(0.09)	(0.08)	(0.10)
Ref years post PhD			0.05***	0.04***	0.05***
			(0.01)	(0.01)	(0.01)
Ref years post PhD squared			-0.00***	-0.00***	-0.00***
			(0.00)	(0.00)	(0.00)
Ref PhD institution group=2			-0.34***	-0.34***	-0.37***
			(0.10)	(0.10)	(0.11)
Ref PhD institution group=3			-0.23**	-0.25**	-0.31***
			(0.10)	(0.10)	(0.11)
Ref PhD institution group=4			-0.01	0.15	0.17
			(0.19)	(0.28)	(0.29)
Ref PhD institution group=5			0.33	0.28	
			(0.22)	(0.22)	
Ref institution FE=2			-0.16*	-0.21**	-0.28***
			(0.08)	(0.09)	(0.09)
Ref institution FE=3			0.02	0.08	0.08
			(0.11)	(0.11)	(0.11)
Ref institution FE=4			-0.06	-0.12	-0.09
			(0.12)	(0.12)	(0.13)
Ref institution FE=5			-0.08	-0.11	-0.08
			(0.21)	(0.23)	(0.24)
Ref no PhD DV			-0.69***	-0.61***	
			(0.12)	(0.14)	
Ref # solo pub			-0.01***	-0.01***	-0.01***
			(0.00)	(0.00)	(0.00)
Ref # joint pub			-0.00**	-0.00	-0.00
			(0.00)	(0.00)	(0.00)
Ref no pub DV			0.06	0.21	
			(0.33)	(0.48)	
Ref # citations			0.00***	0.00	-0.00
			(0.00)	(0.00)	(0.00)
Ref # citations squared			-0.00	0.00	0.00
			(0.00)	(0.00)	(0.00)

Continued on next page

Table A.4 – continued from previous page

	(1)	(2)	(3)	(4)	(5)
Ref no citation DV			-0.39***	-0.28**	
			(0.10)	(0.11)	
Within			0.17***	0.16**	0.13*
			(0.07)	(0.06)	(0.07)
Ref missing field DV			-0.41*	-0.64**	
			(0.23)	(0.27)	
Ref rank FE=2				0.40*	0.81***
				(0.22)	(0.25)
Ref rank FE=3				0.12	0.41*
				(0.20)	(0.23)
Ref rank FE=4				0.04	0.31
				(0.22)	(0.25)
Ref rank FE=5				0.44*	0.71***
				(0.23)	(0.26)
Ref missing rank DV				0.02	
				(0.40)	
Constant	3.77***	3.10***	3.00***	3.00***	2.69***
	(0.04)	(0.28)	(0.30)	(0.34)	(0.35)
N	853	853	853	853	699
Year FE	X	X	X	X	X
Paper controls		X	X	X	X
Referee controls			X	X	X
Rank fixed effects				X	X
Complete information					X

Note: This table reports the estimated gender gap in confidence, as shown in Table 2 and showing all covariates. The regression specification is given by equation (1). The number of publications, citations, and years post-PhD are imputed to average if we could not find this information about the reviewer (31, 1, 30 reviewer-year pairs, respectively). Column (5) includes only the reviewer-paper pairs for which we have a complete set of information. Field dummies exclude Agricultural and Natural Resource Economics, the omitted variable. Reviewer PhD institution groups are as listed in Panel B of Table A.2, with Group 1 "Harvard, MIT, Yale, Northwestern, Columbia, Penn" as the omitted variable. Reviewer institution groups are: Group 2 "other academia Ireland", Group 3 "ESRI", Group 4 "other outside academia Ireland", Group 5 "other outside Ireland", with Group 1 "UCD and TCD" as the omitted variable. Reviewer ranks are as listed in Panel C of Table A.2, with rank 1 "Economist, PhD student" as the omitted variable. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5- Gender gap in confidence: the role of rank

	Reviewers' confidence score {1;5}			
	(1)	(2)	(3)	(4)
Female	0.07	0.10	**0.02**	0.05
	(0.10)	(0.11)	(0.11)	(0.12)
Female x Senior	-0.03	-0.11	-0.10	-0.04
	(0.15)	(0.16)	(0.16)	(0.18)
Female x Prof	0.48***	0.48***	0.70***	0.64***
	(0.14)	(0.15)	(0.17)	(0.18)
Senior	0.12	0.14	-0.05	-0.09
	(0.10)	(0.09)	(0.10)	(0.11)
Prof	0.48***	0.50***	0.12	0.12
	(0.08)	(0.08)	(0.12)	(0.13)
N	853	853	853	699
Year FE	X	X	X	X
Paper controls		X	X	X
Referee controls			X	X
Complete information				X
Gender gap if Senior	0.04	-0.01	**-0.07**	0.01
	[0.71]	[0.96]	[0.55]	[0.96]
Gender gap if Prof	0.55	0.59	**0.72**	0.69
	[0.00]	[0.00]	[0.00]	[0.00]

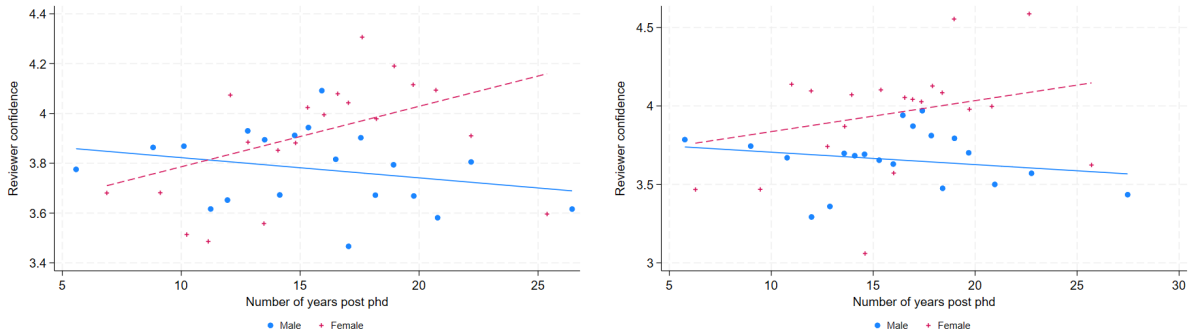
Note: This table reports the estimated gender gap in confidence by reviewer's rank. The regression specification is given by equation (1), interacting the female dummy with a Professor and Senior dummies. Senior includes rank 4, Professor is defined as rank 5. The omitted category is Junior: rank ≤ 3. Robust standard errors in parentheses; P-values in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Coefficients in bold are plotted in Figure 2. Robustness check excluding influential reviewer is in Table D.1.

Table A.6- Gender gap in confidence: the role of expertise by rank

	Reviewers' confidence score {1;5}
	(1)
Female	0.16
	(0.11)
Prof	0.27**
	(0.12)
Female × Prof	0.66***
	(0.17)
Within	0.39***
	(0.09)
Female × Within	-0.45***
	(0.16)
Prof × Within	-0.25*
	(0.14)
Female × Prof × Within	-0.06
	(0.29)
N	853
Year FE	X
Paper controls	X
Referee controls	X
Gender gap if Within=1 & Prof=0	-0.29
	[0.02]
Gender gap if Within=0 & Prof=1	**0.82**
	[0.00]
Gender gap if Within=1 & Prof=1	**0.31**
	[0.19]
Expertise effect if Female=1 & Prof=0	-0.06
	[0.67]
Expertise effect if Female=0 & Prof=1	**0.14**
	[0.23]
Expertise effect if Female=1 & Prof=1	**-0.37**
	[0.09]
Gender Diff in Expertise Effect if Prof=1	**-0.51**
	[0.04]
Gender Diff in Expertise Effect if Prof=0	-0.45
	[0.00]

Note: This table reports the estimated gender gap in confidence by the reviewers' rank and expertise. The regression specification is a triple interaction: female dummy, dummy equal to one if the reviewer is a Professor: rank 5, dummy equal to one if the paper is within the reviewer's field of expertise: Within. The paper is considered within the field of expertise if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. The controls are the same as in equation (1). Robust standard errors in parentheses; P-values in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Coefficients in bold are plotted in Figure 3.

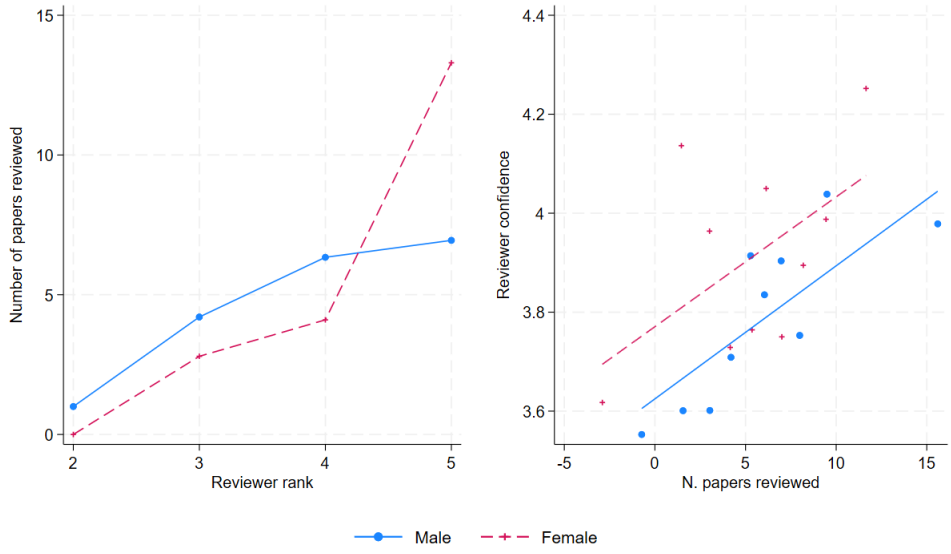
Figure A.2- The role of experience in he profession



(a) Full sample **(b)** Papers outside the field

Notes: Binned scatter plot displaying the correlation between reviewers' confidence and the years spent in the profession after the PhD, net of paper and reviewer controls, as in specification (1), separately for men and women. Panel (a) includes all reviewers; panel (b) only reviewers evaluating papers outside the field of expertise. Within expertise is a dummy equal to 1 if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar.

Figure A.3- Confidence and reviewing experience



Notes: Binned scatter plot displaying the correlation between rank and N. of papers reviewed (left), and the correlation between the N. of papers reviewed and the reviewers' confidence score (right). Rank 2 is Ph.D. student, Post-doc (or Research Fellow) or equivalent, 3 is Assistant Professor or equivalent, 4 is Associate Professor or equivalent, and 5 is Professor or equivalent.

Table A.7- Gender gap in declared confidence: the role of reviewing experience

	Reviewers' confidence score {1;5}	
	(1)	(2)
Female	0.26***	0.47***
	(0.09)	(0.11)
Female × Rookie	-0.19	-0.21
	(0.15)	(0.18)
Female × Within		-0.55***
		(0.16)
Female × Within × Rookie		0.05
		(0.28)
Within	0.16**	0.30***
	(0.06)	(0.09)
Rookie	-0.14	-0.17
	(0.09)	(0.11)
Within × Rookie		0.03
		(0.14)
N	853	853
Year FE	X	X
Paper controls	X	X
Referee controls	X	X
Gender gap if Experienced and within expertise		-0.08
		[0.58]
Gender gap if Rookie and outside expertise		0.26
		[0.08]
Gender gap if Rookie and within expertise		-0.24
		[0.20]
Expertise Effect if Experienced and Female		-0.25
		[0.08]
Expertise Effect if Rookie and Male		0.34
		[0.01]
Expertise Effect if Rookie and Female		-0.16
		[0.43]
Gender Diff in Expertise Effect if Rookie		-0.49
		[0.03]
Gender Diff in Expertise Effect if Experienced		-0.55
		[0.00]

Note: This table reports the estimated gender gap in confidence by the reviewers' experience reviewing for the conference and expertise. The regression specification is a triple interaction: female dummy, dummy equal to one if the reviewer has never reviewed a submission for the conference before: Rookie, dummy equal to one if the paper is within the reviewer's field of expertise: Within. The paper is considered within the field of expertise if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. The controls are the same as in equation (1). Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8- Gender gap in confidence: the role of expertise

	Reviewers' confidence score {1;5}			
	(1)	(2)	(3)	(4)
Female	0.38***	0.42***	**0.36***	0.41***
	(0.09)	(0.10)	(0.09)	(0.10)
Within	0.33***	0.31***	**0.30***	0.28***
	(0.07)	(0.08)	(0.07)	(0.08)
Female x Within	-0.48***	-0.55***	**-0.48***	-0.52***
	(0.14)	(0.15)	(0.14)	(0.15)
N	853	853	853	699
Year FE	X	X	X	X
Paper controls		X	X	X
Referee controls			X	X
Complete information				X
Gender gap if Within=1	-0.10	-0.13	**-0.13**	-0.11
	[0.34]	[0.26]	[0.26]	[0.38]
Expertise effect if Female=1	-0.15	-0.24	**-0.18**	-0.24
	[0.19]	[0.06]	[0.12]	[0.07]

Note: This table reports the estimated gender gap in confidence by reviewer's expertise. The regression specification is given by equation (1), interacting the female dummy with a Within dummy equal to one if the paper is within the reviewer's field of expertise. The paper is considered within the field of expertise if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. Robust standard errors in parentheses; P-values in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robustness check excluding influential reviewer is in Table D.2. Coefficients in bold are plotted in Figure 4.

Table A.9- Female Representation by Field in Economics

Primary Field	Mapped NEP Field	Female Share (%)
International Economics	NEP-INT: International Trade	25.9
Health, Education, and Welfare Economics	NEP-HEA: Health Economics	33.4
Financial Economics	NEP-FIN: Finance	13.2
Public Economics	NEP-PBE: Public Economics	22.4
Microeconomics	NEP-MIC: Microeconomics	14.0
Labour/Demographic Economics	NEP-LAB: Labour Economics	31.0
	NEP-DEM: Demographic Economics	42.2
Macroeconomics	NEP-MAC: Macroeconomics	19.2
Agricultural and Natural Resource Economics	NEP-AGR: Agricultural Economics	27.8
Regional/Real Estate/Transport Economics	NEP-URE: Urban and Real Estate Economics	26.4
	NEP-TRE: Transport Economics	26.2
Economic Development and Growth	NEP-DEV: Development	30.2
Industrial Organisation	NEP-IND: Industrial Organization	19.3
History of Economic Thought	NEP-HPE: History and Philosophy of Economics	16.3
Economic History	NEP-HIS: Business, Economic and Financial History	21.4
Teaching Economics	NEP-EDU: Education	31.3
Energy Economics	NEP-ENE: Energy Economics	23.8
Household Finance and Consumption	NEP-MFD: Microfinance	26.9
Economic Systems	NEP-LMA: Labor Markets - Supply, Demand and Wages	28.9
Housing Economics	NEP-URE: Urban and Real Estate Economics	26.4

Note: This table displays the share of female among economists working in each field. Source: Field mapping based on NEP field statistics from IDEAS Repec Website at https://ideas.repec.org/top/female.html#field, accessed 17 Mar 2025.

Table A.10- Gender gap in confidence: the role of the paper's field stereotypicality

	Reviewers' confidence score {1;5}
	(1)
Female	1.07***
	(0.21)
Female Share	8.90
	(11.25)
Female × Female Share	-3.67***
	(0.83)
N	853
Year FE	X
Paper controls	X
Referee controls	X

Note: This table reports the estimated gender gap in confidence by gender stereotypicality of the paper's topic. The regression specification is given by equation (1), interacting the female dummy with the female representation in each field: Female Share ([0;1]), constructed using the statistics regarding the share of women working in each field in Economics provided by IDEAS. The categorization can be found in Table A.9 in Appendix. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robustness check excluding influential reviewer is in Table D.3.

Table A.11- Gender gap in confidence: authors' gender

	Reviewers' confidence score {1;5}			
	Male submitter		Prop. male authors	
	(1)	(2)	(3)	(4)
Female	**0.08**	**0.02**	0.16	0.02
	(0.14)	(0.17)	(0.17)	(0.22)
Male Authors	-0.14	-0.30***	-0.06	-0.37**
	(0.09)	(0.11)	(0.11)	(0.14)
Male Authors × Female	0.14	0.43**	0.02	0.41
	(0.16)	(0.20)	(0.19)	(0.26)
Within	0.18**	0.02	0.18**	-0.20
	(0.07)	(0.14)	(0.07)	(0.18)
Within × Male Authors		0.39**		0.68***
		(0.16)		(0.21)
Within × Female		0.15		0.28
		(0.26)		(0.32)
Within × Male Authors × Female		-0.79**		-0.96**
		(0.31)		(0.39)
N	666	666	666	666
Gender gap if Male Authors	**0.22**		0.18	
	[0.04]		[0.09]	
Expertise Effect if Male Author and Male Reviewer		**0.40**		0.48
		[0.00]		[0.00]
Expertise Effect if Female Author and Female Reviewer		**0.17**		0.08
		[0.45]		[0.77]
Expertise Effect if Male Author and Female Reviewer		**-0.23**		-0.20
		[0.14]		[0.20]
Gender Diff in Expertise Effect if Male Author		**-0.64**		-0.68
		[0.00]		[0.00]
Gender Diff in Expertise Effect if Female Author		**0.15**		0.28
		[0.57]		[0.38]
Year FE		X		X
Paper controls		X		X
Referee controls		X		X

Note: This table reports the estimated gender gap in confidence by authors' gender. The regression specification in columns 1 and 3 is given by equation (1), interacting the female dummy with a dummy equal to one if the submitter of the paper is male (column 1), and the proportion of authors who are male (column 3). In columns 2 and 4, the specification is given by a triple interaction: female dummy, dummy equal to one if the submitter of the paper is male (column 2) and the proportion of authors who are male (column 4), and a Within dummy equal to one for papers within the reviewer's field of expertise. The paper is considered within the field of expertise if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. Controls are the same as in equation (1). The sample includes only non-blind submissions for which the identity of the authors/submitters is known. Robust standard errors in parentheses; P-values in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robustness check excluding influential reviewer is in Table D.4. Coefficients in bold are plotted in Figure 6.

Table A.12- Gender gap in confidence: the role of lower-ranked institutions

	Reviewers' confidence score {1;5}	
	(1)	(2)
Female	**-0.35***	-0.26
	(0.21)	(0.25)
Lower-ranked	0.13	0.14
	(0.13)	(0.15)
Female × Lower-ranked	0.42*	0.49*
	(0.23)	(0.28)
Within	0.18***	**0.32****
	(0.07)	(0.16)
Female × Within		-0.23
		(0.27)
Within × Lower-ranked		-0.04
		(0.18)
Female × Within × Lower-ranked		-0.18
		(0.31)
N	690	690
Gender gap if Lower-ranked=1	**0.07**	
	[0.41]	
Lower-ranked effect if Female=1	0.55	
	[0.00]	
Gender gap if Lower-ranked=1 and Within=0		0.23
		[0.03]
Gender gap if Lower-ranked=0 and Within=1		-0.49
		[0.04]
Expertise Effect if Lower-ranked=1 and Female=0		**0.28**
		[0.00]
Expertise Effect if Lower-ranked=0 and Female=1		**0.09**
		[0.69]
Expertise Effect if Lower-ranked=1 and Female=1		**-0.13**
		[0.37]
Gender Diff in Expertise Effect if Lower-ranked=1		**-0.41**
		[0.01]
Gender Diff in Expertise Effect if Lower-ranked=0		**-0.23**
		[0.39]
Year FE	X	X
Paper controls	X	X
Referee controls	X	X

Note: This table reports the estimated gender gap in confidence by rank of reviewer PhD institution, excluding organizers. The regression specification in column 1 is given by equation (1), interacting the female dummy with a dummy equal to one for reviewers with a PhD from a lower-ranked institution (not Harvard, MIT, Yale, Northwestern, Columbia, Penn, LSE, UCL, Oxford, Warwick). In column 2, the specification is given by a triple interaction: female dummy, dummy equal to one for reviewers from lower-ranked institutions, and a Within dummy equal to one for papers within the reviewer's field of expertise. The paper is considered within the field of expertise if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. Controls are the same as in equation (1). Robust standard errors in parentheses; P-values in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robustness check excluding influential reviewer is in Table D.5. Coefficients in bold are plotted in Figure 7.

Table A.13- Gender gap in confidence: the role of being familiar with the organisers

	Reviewers' confidence score {1;5}	
	(1)	(2)
Female	**0.17***	0.40***
	(0.09)	(0.12)
Colleague	0.07	-0.02
	(0.09)	(0.11)
Female × Colleague	-0.38**	-0.44**
	(0.15)	(0.20)
Within	0.14**	**0.24*****
	(0.07)	(0.09)
Female × Within		-0.53***
		(0.18)
Within × Colleague		0.15
		(0.17)
Female × Within × Colleague		0.10
		(0.30)
N	690	690
Gender gap if Colleague=1	**-0.21**	
	[0.09]	
Colleague effect if Female=1	-0.31	
	[0.01]	
Gender gap if Colleague=1 and Within=1		-0.47
		[0.02]
Expertise Effect if Colleague=1 and Female=0		**0.38**
		[0.01]
Expertise Effect if Colleague=0 and Female=1		**-0.29**
		[0.07]
Expertise Effect if Colleague=1 and Female=1		**-0.04**
		[0.86]
Gender Diff in Expertise Effect if Colleague=1		**-0.42**
		[0.09]
Gender Diff in Expertise Effect if Colleague=0		**-0.53**
		[0.00]
Year FE	X	X
Paper controls	X	X
Referee controls	X	X

Note: This table reports the estimated gender gap in confidence by reviewers' familiarity with the organizers. The regression specification is given by equation (1), interacting the female dummy with a dummy equal to one for reviewers working in the same institution as the organizers: column 1. In column 2, the specification is given by a triple interaction: female dummy, dummy equal to one for reviewers working in the same institution as the organizers, dummy equal to one for papers within the reviewer's field of expertise: Within. The paper is considered within the field of expertise if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. Controls are the same as in equation (1). The sample includes only reviewers who are not organizers. Robust standard errors in parentheses; P-values in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robustness check excluding influential reviewer is in Table D.6. Coefficients in bold are plotted in Figure 8.

Table A.14- Reviewers' Commenting Behavior

	Pr(leaving a comment)				Effort Score {1;5}			
Subsample:	All	All	Junior+Senior	Professor	All	All	Junior+Senior	Professor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.13***	-0.18***	-0.13**	-0.32**	-0.38***	-0.54***	-0.50***	-0.67
	(0.03)	(0.04)	(0.05)	(0.13)	(0.12)	(0.14)	(0.18)	(0.46)
Within	-0.09***	-0.13***	-0.12**	-0.09	-0.25**	-0.37***	-0.44***	-0.14
	(0.03)	(0.04)	(0.05)	(0.06)	(0.11)	(0.13)	(0.17)	(0.21)
Female × Within		0.13*	0.07	0.41**		0.41*	0.36	1.04*
		(0.07)	(0.08)	(0.18)		(0.25)	(0.30)	(0.62)
N	701	701	491	210	701	701	491	210
Year FE	X	X	X	X	X	X	X	X
Paper controls	X	X	X	X	X	X	X	X
Referee controls	X	X	X	X	X	X	X	X
Gender gap if Within=1		-0.05	-0.05	0.09		-0.13	-0.14	0.37
		[0.38]	[0.44]	[0.46]		[0.52]	[0.56]	[0.38]

Note: This table reports the estimated gender gap in commenting behavior by reviewer's expertise and rank. The outcome variables are the probability of leaving a comment in columns 1-4 and an effort score in columns 5-8, which is a score produced by Generative AI when asked to classify comments based on the level of effort put into the review on a scale from 1-5. No comment or comments equal to "na" are given a score of zero. The regression specification is given by equation (1). Within is a dummy equal to one if the paper is within the reviewer's field of expertise. The paper is considered within the field of expertise if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. Columns 1, 2, 5, 6 include the full sample of reviewers excluding organizers, while columns 3-4 and 7-8 split the sample in Junior+Senior and Professor, respectively. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.15- The gender influence gap

	Organizers' acceptance decision {0;1}			
	(1)	(2)	(3)	(4)
Referee's acceptance score	0.26***	0.25***		
	(0.01)	(0.01)		
Female	-0.07**	-0.19***	-0.01	0.02
	(0.03)	(0.07)	(0.04)	(0.04)
Female × Referee's acceptance score		0.03*		
		(0.02)		
Referee's acceptance score=3			0.25***	0.31***
			(0.06)	(0.07)
Referee's acceptance score=4			0.67***	0.68***
			(0.04)	(0.05)
Referee's acceptance score=5			0.88***	0.89***
			(0.04)	(0.04)
Referee's acceptance score=3 × Female			-0.26***	-0.33***
			(0.07)	(0.08)
Referee's acceptance score=4 × Female			0.00	-0.00
			(0.07)	(0.07)
Referee's acceptance score=5 × Female			-0.04	-0.05
			(0.05)	(0.05)
N	853	853	853	699
Year FE	X	X	X	X
Paper controls	X	X	X	X
Referee controls	X	X	X	X
Complete information				X
Gender gap if Referee's acceptance score=3			-0.27	-0.32
			[0.00]	[0.00]
Gender gap if Referee's acceptance score=4			-0.01	0.16
			[0.92]	[0.79]
Gender gap if Referee's acceptance score=5			-0.05	-0.03
			[0.11]	[0.34]

Note: This table reports the estimated gender gap in influence. The outcome variable is the organizers' acceptance decision ({0,1}). The regression specification is given by equation (2). Columns 1-2 use the continuous reviewer acceptance score (1-5), columns 3-4 use the categorical values of reviewer acceptance score, with scores ≤ 2 as the omitted category. The number of publications, citations, and years post-PhD are imputed to average if we could not find this information about the reviewer (31, 1, 30 reviewer-year pairs, respectively). Column (4) includes only the reviewer-paper pairs for which we have a complete set of information. Robust standard errors in parentheses; P-values in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.16- Summary: The gender influence gap by declared confidence and rank

	Organizers' acceptance decision {0;1}
	(1)
Panel A: Gender gap if low declared confidence	
Referee's acceptance score=3	-0.20
	[0.02]
Referee's acceptance score=4	0.17
	[0.03]
Referee's acceptance score=5	-0.10
	[0.12]
Panel B: Gender gap if average declared confidence	
Referee's acceptance score$\leq$2	0.14
	[0.01]
Referee's acceptance score=3	-0.35
	[0.00]
Referee's acceptance score=4	-0.14
	[0.11]
Referee's acceptance score=5	-0.03
	[0.56]
Panel C: Gender gap if high declared confidence	
Referee's acceptance score$\leq$2	-0.11
	[0.21]
Referee's acceptance score=3	-0.27
	[0.08]
Referee's acceptance score=4	-0.34
	[0.05]
Referee's acceptance score=5	-0.06
	[0.22]
Panel D: Gender gap if Junior+Senior	
Referee's acceptance score=3	-0.28
	[0.00]
Referee's acceptance score=4	0.04
	[0.54]
Referee's acceptance score=5	0.00
	[0.98]
Panel E: Gender gap if Prof	
Referee's acceptance score$\leq$2	-0.09
	[0.23]
Referee's acceptance score=3	-0.26
	[0.02]
Referee's acceptance score=4	-0.25
	[0.04]
Referee's acceptance score=5	-0.17
	[0.00]

Note: Gender influence gap for different acceptance scores by declared confidence levels and ranks. Estimates from Table A.17 in Panels A-C and Table A.18 in Panels D and E. P-values in square brackets.

Table A.17- The gender influence gap by declared confidence

	Organizers' acceptance decision {0;1}
	(1)
Referee's acceptance score=3	0.19**
	(0.08)
Referee's acceptance score=4	0.54***
	(0.06)
Referee's acceptance score=5	0.89***
	(0.06)
Female	0.01
	(0.05)
Referee's acceptance score=3 × Female	-0.21**
	(0.09)
Referee's acceptance score=4 × Female	0.16*
	(0.09)
Referee's acceptance score=5 × Female	-0.11
	(0.09)
Average Confidence	-0.07
	(0.05)
High Confidence	-0.00
	(0.09)
Referee's acceptance score=3 × Average Confidence	0.18
	(0.12)
Referee's acceptance score=3 × High Confidence	0.01
	(0.19)
Referee's acceptance score=4 × Average Confidence	0.28***
	(0.08)
Referee's acceptance score=4 × High Confidence	0.26**
	(0.12)
Referee's acceptance score=5 × Average Confidence	0.05
	(0.06)
Referee's acceptance score=5 × High Confidence	-0.05
	(0.10)
Female × Average Confidence	0.13*
	(0.07)
Female × High Confidence	-0.12
	(0.11)
Referee's acceptance score=3 × Female × Average Confidence	-0.28*
	(0.15)
Referee's acceptance score=3 × Female × High Confidence	0.05
	(0.21)
Referee's acceptance score=4 × Female × Average Confidence	-0.44***
	(0.13)
Referee's acceptance score=4 × Female × High Confidence	-0.38*
	(0.21)
Referee's acceptance score=5 × Female × Average Confidence	-0.06
	(0.11)
Referee's acceptance score=5 × Female × High Confidence	0.17
	(0.13)
N	853
Year FE	X
Paper controls	X
Referee controls	X

Note: This table reports the estimated gender gap in influence by reviewer's declared confidence. The outcome variable is the organizers' acceptance decision ($\{0,1\}$). The regression specification is given by equation (2), interacting the female x score interaction with a dummy equal to one if the reviewer declares average confidence, and a dummy equal to one if the reviewer declares high confidence. Low, average, and high confidence correspond to confidence scores less than or equal to 3, equal to 4, and equal to 5, respectively. The categorical values of reviewer acceptance score are used, with scores ≤ 2 as the omitted category. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.18- The gender influence gap by rank

	Organizers' acceptance decision {0;1}
	(1)
Female	0.04
	(0.05)
Referee's acceptance score=3	0.24***
	(0.06)
Referee's acceptance score=4	0.66***
	(0.05)
Referee's acceptance score=5	0.85***
	(0.04)
Female × Referee's acceptance score=3	-0.32***
	(0.08)
Female × Referee's acceptance score=4	-0.00
	(0.08)
Female × Referee's acceptance score=5	-0.04
	(0.06)
Prof	0.02
	(0.08)
Female × Prof	-0.13
	(0.09)
Referee's acceptance score=3 × Prof	0.01
	(0.13)
Referee's acceptance score=4 × Prof	-0.01
	(0.10)
Referee's acceptance score=5 × Prof	0.05
	(0.08)
Female × Referee's acceptance score=3 × Prof	0.16
	(0.16)
Female × Referee's acceptance score=4 × Prof	-0.16
	(0.16)
Female × Referee's acceptance score=5 × Prof	-0.04
	(0.11)
N	853
Year FE	X
Paper controls	X
Referee controls	X

Note: This table reports the estimated gender gap in influence by reviewer's rank. The outcome variable is the organizers' acceptance decision ({0,1}). The regression specification is given by equation (2), interacting the female x score interaction with a Professor dummy. Professor is a dummy equal to one for rank 5. The categorical values of reviewer acceptance score are used, with scores ≤ 2 as the omitted category. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

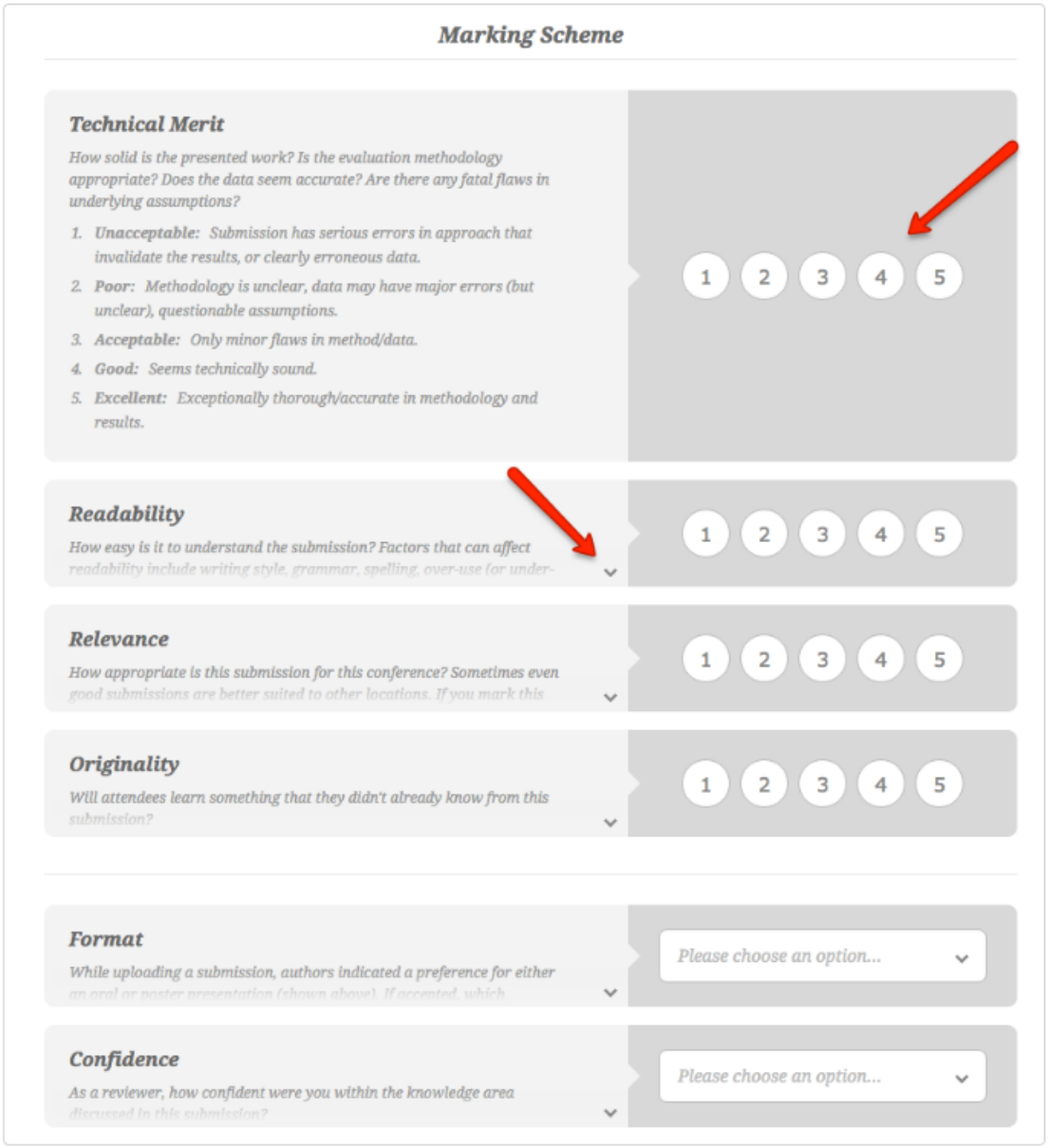
Table A.19- Gender gap in reviewers' leniency

	Acceptance Score (1)	**Acceptance Score≤2** (2)	**Acceptance Score=3** (3)	**Acceptance Score=4** (4)	**Acceptance Score=5** (5)	**Average Score** (6)
Panel A: All reviewers						
Female	0.13	-0.02	-0.02	0.02	0.01	0.09
	(0.09)	(0.03)	(0.03)	(0.04)	(0.04)	(0.07)
N	853	853	853	853	853	853
Panel B: Professors only						
Female	0.24	0.08	-0.19**	-0.09	0.20	0.16
	(0.22)	(0.05)	(0.08)	(0.12)	(0.13)	(0.17)
N	252	252	252	252	252	252
Year FE	X	X	X	X	X	X
Paper controls	X	X	X	X	X	X
Referee controls	X	X	X	X	X	X

Note: This table displays the estimated gender gap in leniency: regression of reviewers' acceptance and average scores on a female dummy, year FEs, and reviewer and paper controls, as specified in equation (1). The outcome variable is the reviewers' acceptance score ({1,5}) in column 1; a dummy equal to one for reviewers' acceptance score ≤ 2, equal to 3, 4, and 5 in columns 2-5, respectively; and the reviewers' average score in column 6: average of technical merit, readability, originality, and relevance scores. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

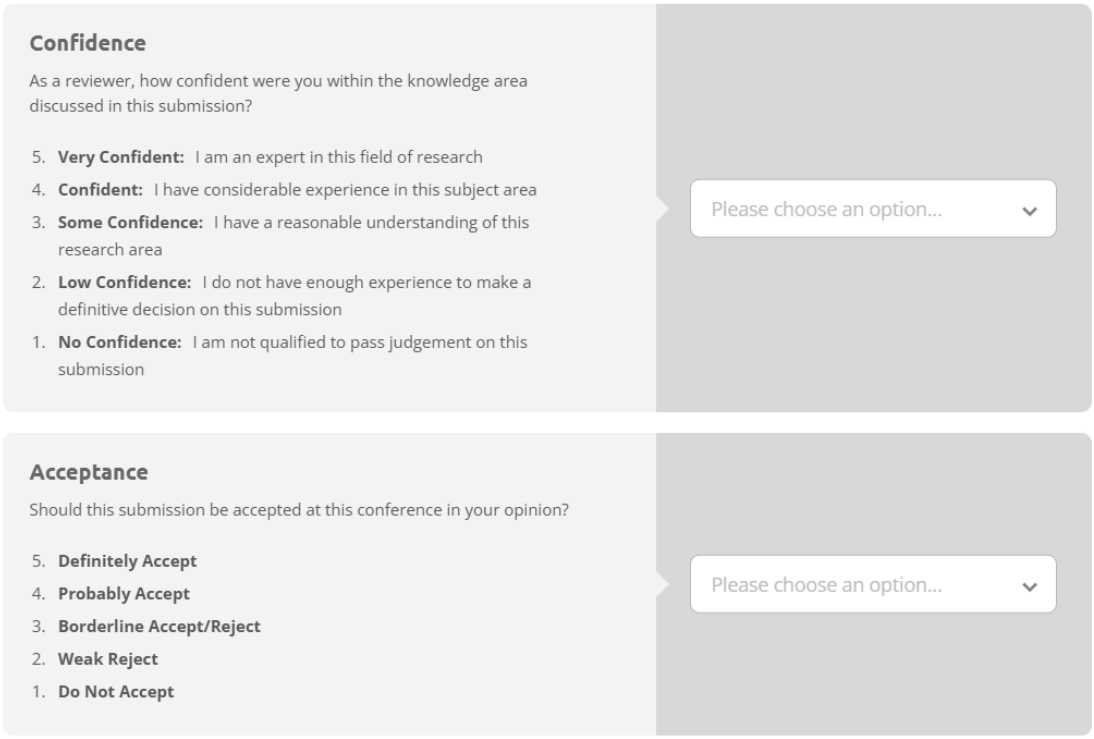
B Marking Scheme

Figure B.1- Marking Scheme



Notes: Screenshot of reviewers' marking scheme for each paper. Source: Reviewers' Pack Guide.

Figure B.2- Confidence and acceptance questions



Notes: Screenshot of acceptance and confidence questions. Source: Reviewers' Pack Guide.

C LLM Measure of Expertise

The measure of expertise used in the main text is based on self-reported data: we classify the reviewer to be an expert for the paper reviewed if the paper's primary field, as declared by the author, matches one of the reviewer's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar.

As a robustness check, we use LLM to classify reviewers into expert or non-expert as described below. First, using Generative AI, we classify each submitted paper into two JEL codes/letters based on the paper title. We use the following prompt:

> Note the following topics/fields in economics. I will subsequently ask you to classify a list of papers into two of these fields based on the paper's title.
>
> - A General Economics and Teaching
> - B History of Economic Thought, Methodology, and Heterodox Approaches
> - C Mathematical and Quantitative Methods

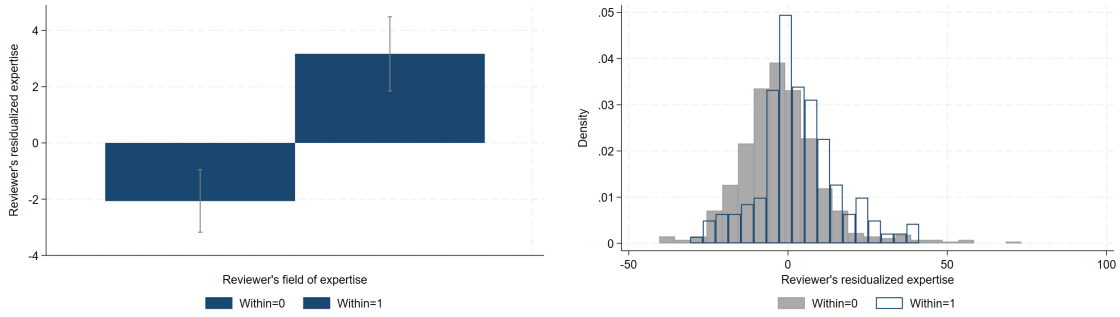
- ...

> I now give you the paper titles. Could you classify these using the above, and output a table with the following 3 columns: paper title, classification letter and label 1, classification letter and label 2, that I can copy/paste to excel?

Having collected all reviewers' journal publications, we then repeat the above for all reviewers' publications to also obtain two JEL codes for each publication. We next calculate each reviewer's publication count per JEL code up to the year before the conference. We define **reviewer expertise** as the **sum** of their publications in both JEL codes of the submitted paper, up to year before conference. We refer to this measure as "objective reviewers' expertise" to differentiate it from the self-reported measure of expertise used in the main text.

As a first sanity check, Figure C.1 displays the distribution of the residuals of objective reviewer's expertise obtained, net of reviewer characteristics and year-fixed effects, by subjective field of expertise (measure used in the main analysis). The correlation between the self-reported measure of expertise used in the main text, and the objective measure of expertise measured using the LLM classification of paper titles is strong. In particular, we can see that papers classified as outside the field of expertise based on self-reported data have a negative average objective expertise, while those classified as within the field of expertise display a positive average objective expertise score (panel a, difference in means p-value=0.00). Moreover, panel (b) shows that the distributions of objective reviewers' expertise scores by subjective expertise, while partially overlapping, are significantly different (Kolmogorov-Smirnov test of difference in distribution p-value=0.00).

Figure C.1- Distribution of expertise by field of expertise

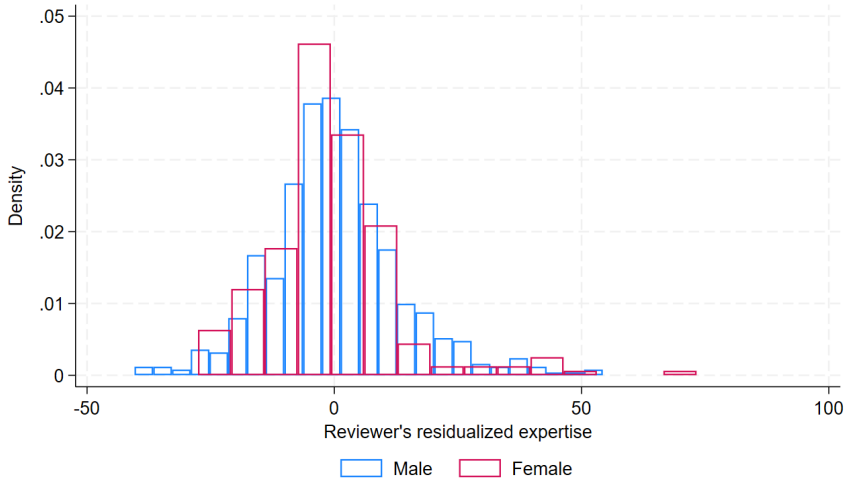


(a) Difference in mean **(b)** Difference in distribution

Notes: Panel (a) displays the average of the residuals of the objective reviewers' expertise score, net of reviewer characteristics and year-fixed effects, by reviewers' subjective field of expertise (measure used in the main analysis). Panel (b) shows the histogram of these residuals by reviewers' subjective field of expertise. Reviewer characteristics include affiliation FEs, N. years post Ph.D. (second order polynomial), Ph.D. institution group FEs, rank FEs, N. of publications (solo and co-authored separately) and N. of citations at the time of the conference (second order polynomial).

Secondly, Figure C.2 displays the distribution of residuals of male and female objective reviewers' expertise, net of reviewer characteristics and year fixed effects. On average, male and female reviewers do not differ in expertise when it is measured using the LLM classification of paper titles, again confirming what we find using the subjective measure of expertise (Table 1).

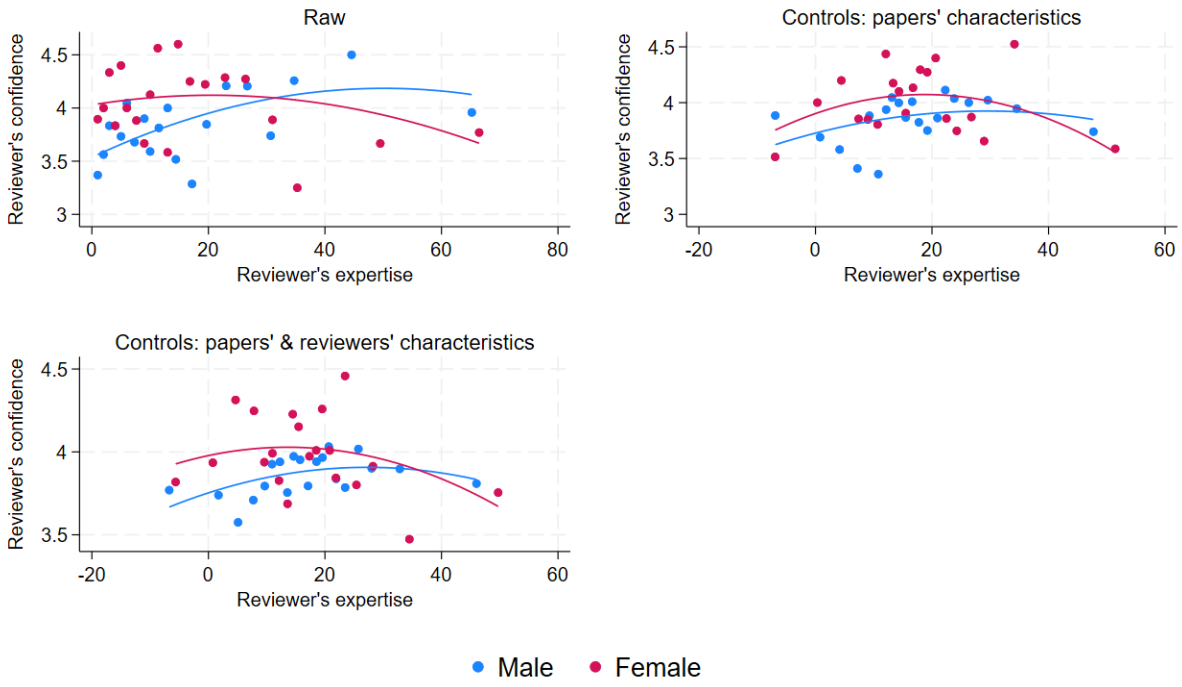
Figure C.2- Distribution of expertise



Notes: Histogram of residuals from a regression of the reviewers' expertise on reviewer characteristics including: affiliation FEs, N. years post Ph.D. (second order polynomial), Ph.D. institution group FEs, rank FEs, N. of publications (solo and co-authored separately) and N. of citations at the time of the conference (second order polynomial).

Finally, Figure C.3 plots the correlation of reviewer expertise and confidence. The top left panel displays the raw correlation, the top right panel the correlation controlling for paper characteristics, while the bottom left panel the correlation controlling for paper and reviewer characteristics. Consistent with earlier findings, we observe that male reviewers' declared confidence is higher for papers characterized by higher values of reviewers' objective expertise—within the field of expertise, while female reviewers' declared confidence is higher for papers characterized by lower values of reviewers' objective expertise—outside the field of expertise.

Figure C.3- Correlation of reviewer expertise and confidence



Notes: Binned scatter plots displaying the correlation between reviewers' confidence scores and reviewers' objective expertise. The top left panel displays the raw correlation, the top right panel the correlation controlling for paper characteristics, while the bottom left panel the correlation controlling for paper and reviewer characteristics. Reviewer characteristics include affiliation FEs, N. years post Ph.D. (second order polynomial), Ph.D. institution group FEs, rank FEs, N. of publications (solo and co-authored separately) and N. of citations at the time of the conference (second order polynomial). The paper-specific characteristics we control for are those specified in equation (2).

D Robustness Analyses Excluding Influential Reviewer

In this section, we present the results of a robustness check in which we test whether our results are driven by particular reviewers.

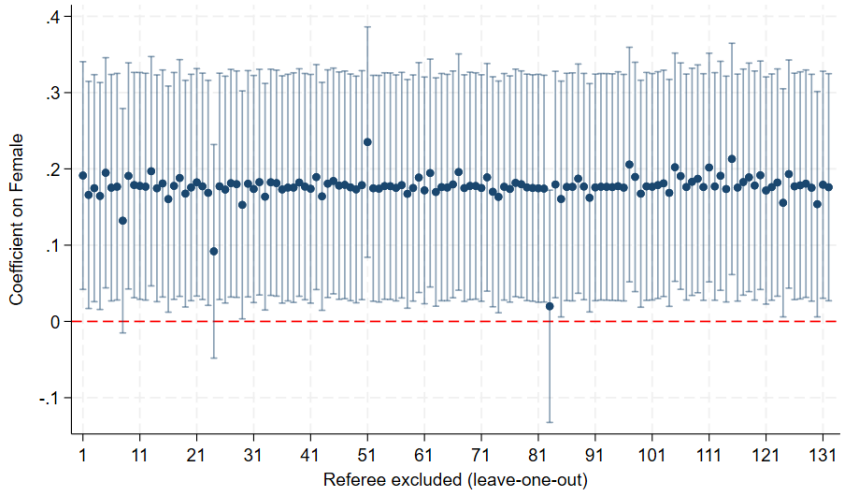
Figure D.1 shows the gender gap in confidence (Female dummy coefficient) obtained by esti-

mating equation (1) leaving out one reviewer at a time. The baseline Female coefficient is 0.18, as shown in column 4 of Table 2. The estimated average gender gap is robust to the exclusion of all the reviewers, with the exception of one. Indeed, the average gender gap turns to zero when we exclude reviewer 83, a female professor who accounts for 54% of papers reviewed by female professors.

This is due to a weighting issue. As discussed in the main text, the positive average gender gap in the full sample is driven by professors. Excluding reviewer 83 reduces the estimated gender gap among professors, although it remains positive and statistically significant (0.41, $p = 0.03$, compared with 0.72, $p < 0.01$, in the full sample; see Table D.1). It also substantially reduces the size of the professor subsample, thereby lowering its weight in the overall estimate and explaining why the average gender gap falls to zero.

Importantly, Tables D.2- D.6 show that the finding that the gender confidence gap becomes systematically more positive in contexts where women's expertise may be perceived as more uncertain – the main contribution of the paper – remains robust to excluding this reviewer.

Figure D.1- Gender confidence gap leaving out one reviewer at the time



Note: This figure plots the estimated gender gap in confidence, leaving out one reviewer at a time. The regression specification is given by equation (1). The baseline Female coefficient is 0.18, as shown in column 4 of Table 2.

Table D.1- Gender confidence gap, the role of rank, leaving out influential reviewer

	Reviewers' confidence score {1;5}	
Sample:	Full	Selected
	(1)	(2)
Female	0.02	0.03
	(0.11)	(0.11)
Female × Senior	-0.10	-0.24
	(0.16)	(0.17)
Female × Prof	0.70***	0.37*
	(0.17)	(0.22)
N	853	805
Year FE	X	X
Paper controls	X	X
Referee controls	X	X
Rank fixed effects	X	X
Gender gap if Senior	-0.07	-0.21
	[0.55]	[0.11]
Gender gap if Prof	0.72	0.41
	[0.00]	[0.03]

Note: This table reports the estimated gender gap in confidence by reviewer's rank. The regression specification is given by equation (1), interacting the female dummy with Professor and Senior dummies. Senior includes rank 4, Professor is defined as rank 5. The omitted category is Junior: rank ≤ 3. Column 1 presents the same baseline results as in column 3 of Table A.5. As one reviewer accounts for 54% of observations involving female professors, column 2 provides a robustness check excluding this influential reviewer. Robust standard errors in parentheses; p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.2- Gender confidence gap, the role of expertise, leaving out influential reviewer

	Reviewers' confidence score {1;5}	
Sample:	Full	Selected
	(1)	(2)
Female	0.36***	0.17*
	(0.09)	(0.10)
Within	0.30***	0.30***
	(0.07)	(0.08)
Female × Within	-0.48***	-0.40***
	(0.14)	(0.15)
N	853	805
Year FE	X	X
Paper controls	X	X
Referee controls	X	X
Rank fixed effects	X	X
Gender gap if Within=1	-0.13	-0.22
	[0.26]	[0.06]
Expertise effect if Female=1	-0.18	-0.09
	[0.12]	[0.50]

Note: This table reports the estimated gender gap in confidence by reviewer's expertise. The regression specification is given by equation (1), interacting the female dummy with a Within dummy equal to one if the paper is within the reviewer's field of expertise. Column 1 presents the same baseline results as in column 3 of Table A.8. As one reviewer accounts for 54% of observations involving female professors, column 2 provides a robustness check excluding this influential reviewer. Robust standard errors in parentheses; p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3- Gender confidence gap, the field stereotypicality, leaving out influential reviewer

	Reviewers' confidence score {1;5}	
Sample:	Full	Selected
	(1)	(2)
Female	1.07***	1.01***
	(0.21)	(0.24)
Female Share	8.90	11.73
	(11.25)	(12.28)
Female × Female Share	-3.67***	-4.02***
	(0.83)	(0.91)
N	853	805
Year FE	X	X
Paper controls	X	X
Referee controls	X	X
Rank fixed effects	X	X

Note: This table reports the estimated gender gap in confidence by gender stereotypicality of the paper's topic. The regression specification is given by equation (1), interacting the female dummy with the female representation in each field: Female Share ([0;1]), constructed using the statistics regarding the share of women working in each field in Economics provided by IDEAS. Column 1 presents the same baseline results as in Table A.10. As one reviewer accounts for 54% of observations involving female professors, column 2 provides a robustness check excluding this influential reviewer. Robust standard errors in parentheses; p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4- Gender confidence gap, submitter gender, leaving out influential reviewer

	Reviewers' confidence score {1;5}							
	Male submitter				Prop. male authors			
Sample:	Full		Selected		Full		Selected	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.08	0.02	-0.01	-0.13	0.16	0.02	0.04	-0.14
	(0.14)	(0.17)	(0.15)	(0.19)	(0.17)	(0.22)	(0.18)	(0.24)
Male Author	-0.14	-0.30***	-0.12	-0.29***	-0.06	-0.37**	-0.06	-0.36**
	(0.09)	(0.11)	(0.09)	(0.11)	(0.11)	(0.14)	(0.11)	(0.14)
Male Author × Female	0.14	0.43**	-0.01	0.31	0.02	0.41	-0.08	0.31
	(0.16)	(0.20)	(0.18)	(0.23)	(0.19)	(0.26)	(0.21)	(0.29)
Within		0.02		-0.00		-0.20		-0.21
		(0.14)		(0.15)		(0.18)		(0.18)
Within × Female		0.15		0.36		0.28		0.41
		(0.26)		(0.30)		(0.32)		(0.35)
Within × Male Author		0.39**		0.40**		0.68***		0.67***
		(0.16)		(0.16)		(0.21)		(0.21)
Within × Male Author × Female		-0.79**		-0.87**		-0.96**		-0.92**
		(0.31)		(0.35)		(0.39)		(0.42)
N	666	666	619	619	666	666	619	619
Gender gap if Male Author=1	0.22		-0.02		0.18		-0.04	
	[0.04]		[0.87]		[0.09]		[0.75]	
Expertise Effect if Male Authors=1 and Male Reviewer		0.40		0.40		0.48		0.46
		[0.00]		[0.00]		[0.00]		[0.00]
Expertise Effect if Male Authors=0 and Female Reviewer		0.17		0.36		0.08		0.20
		[0.45]		[0.18]		[0.77]		[0.53]
Expertise Effect if Male Authors=1 and Female reviewer		-0.23		-0.11		-0.20		-0.05
		[0.14]		[0.56]		[0.20]		[0.78]
Gender Diff in Expertise Effect if Male Authors=1		-0.64		-0.51		-0.68		-0.52
		[0.00]		[0.02]		[0.00]		[0.02]
Gender Diff in Expertise Effect if Male Authors=0		0.15		0.36		0.28		0.41
		[0.57]		[0.23]		[0.38]		[0.25]
Year FE		X		X		X		X
Paper controls		X		X		X		X
Referee controls		X		X		X		X

Note: This table reports the estimated gender gap in confidence by authors' gender and reviewer's expertise. The regression specifications are given by equation (1) with interactions, using a dummy equal to one if the submitter of the paper is male (columns 1-4) or the proportion of authors who are male (columns 5-8). Columns 1-2, 5-6 present the same baseline results as in Table A.11. As one reviewer accounts for 54% of observations involving female professors, columns 3-4, 7-8 provide a robustness check excluding this influential reviewer. Robust standard errors in parentheses; p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5- Gender confidence gap, lower-ranked institutions, leaving out influential reviewer

	Reviewers' confidence score {1;5}			
Sample:	Full		Selected	
	(1)	(2)	(3)	(4)
Female	-0.35*	-0.26	-0.38*	-0.29
	(0.21)	(0.25)	(0.21)	(0.26)
Lower-ranked	0.13	0.14	0.14	0.15
	(0.13)	(0.15)	(0.13)	(0.15)
Female × Lower-ranked	0.42*	0.49*	0.42*	0.48*
	(0.23)	(0.28)	(0.23)	(0.29)
Within	0.18***	0.32**	0.18***	0.32**
	(0.07)	(0.16)	(0.07)	(0.16)
Female × Within		-0.23		-0.24
		(0.27)		(0.27)
Within × Lower-ranked		-0.04		-0.04
		(0.18)		(0.18)
Female × Within × Lower-ranked		-0.18		-0.17
		(0.31)		(0.32)
N	690	690	678	678
Year FE	X	X	X	X
Paper controls	X	X	X	X
Referee controls	X	X	X	X
Gender gap if Lower-ranked=1	0.07		0.03	
	[0.41]		[0.71]	
Lower-ranked effect if Female=1	0.55		0.56	
	[0.00]		[0.00]	
Gender gap if Lower-ranked=1 and Within=0		0.23		0.19
		[0.03]		[0.07]
Gender gap if Higher-ranked=1 and Within=1		-0.49		-0.53
		[0.04]		[0.03]
Expertise Effect if Higher-ranked=0 and Female=0		0.28		0.28
		[0.00]		[0.00]
Expertise Effect if Higher-ranked=0 and Female=1		-0.13		-0.14
		[0.37]		[0.36]
Expertise Effect if Higher-ranked=1 and Female=1		0.09		0.07
		[0.69]		[0.75]
Gender Diff in Expertise Effect if Higher-ranked=0		-0.41		-0.42
		[0.01]		[0.02]
Gender Diff in Expertise Effect if Higher-ranked=1		-0.23		-0.24
		[0.39]		[0.38]

Note: This table reports the estimated gender gap in confidence by rank of reviewer PhD institution and expertise, excluding organizers. The regression specification in column 1 is given by equation (1), interacting the female dummy with a dummy equal to one for reviewers with a PhD from a lower-ranked institution (not Harvard, MIT, Yale, Northwestern, Columbia, Penn, LSE, UCL, Oxford, Warwick). Columns 1-2 present the same baseline results as in Table A.12. As one reviewer accounts for 54% of observations involving female professors, columns 3-4 provide a robustness check excluding this influential reviewer. Robust standard errors in parentheses; p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.6- Gender confidence gap, familiarity with the organizers, leaving out influential reviewer

	Reviewers' confidence score {1;5}			
Sample:	Full		Selected	
	(1)	(2)	(3)	(4)
Female	0.17*	0.40***	0.13	0.36***
	(0.09)	(0.12)	(0.10)	(0.13)
Colleague	0.07	-0.02	0.07	-0.02
	(0.09)	(0.11)	(0.09)	(0.12)
Female × Colleague	-0.38**	-0.44**	-0.35**	-0.41**
	(0.15)	(0.20)	(0.16)	(0.21)
Within	0.14**	0.24***	0.14**	0.23***
	(0.07)	(0.09)	(0.07)	(0.09)
Female × Within		-0.53***		-0.53***
		(0.18)		(0.19)
Within × Colleague		0.15		0.16
		(0.17)		(0.17)
Female × Within × Colleague		0.10		0.09
		(0.30)		(0.31)
N	690	690	678	678
Year FE	X	X	X	X
Paper controls	X	X	X	X
Referee controls	X	X	X	X
Gender gap if Colleague=1	-0.21		-0.22	
	[0.09]		[0.08]	
Colleague effect if Female=1	-0.31		-0.28	
	[0.01]		[0.03]	
Gender gap if Colleague=1 and Within=1		-0.47		-0.49
		[0.02]		[0.02]
Expertise Effect if Colleague=1 and Female=0		0.38		0.39
		[0.01]		[0.01]
Expertise Effect if Colleague=0 and Female=1		-0.29		-0.30
		[0.07]		[0.08]
Expertise Effect if Colleague=1 and Female=1		-0.04		-0.05
		[0.86]		[0.82]
Gender Diff in Expertise Effect if Colleague=0		-0.53		-0.53
		[0.00]		[0.01]
Gender Diff in Expertise Effect if Colleague=1		-0.42		-0.44
		[0.09]		[0.08]

Note: This table reports the estimated gender gap in confidence by reviewers' familiarity with the organizers and expertise. The regression specification is given by equation (1), interacting the female dummy with a dummy equal to one for reviewers working in the same institution as the organizers. Columns 1-2 present the same baseline results as in Table A.13. As one reviewer accounts for 54% of observations involving female professors, columns 3-4 provide a robustness check excluding this influential reviewer. Robust standard errors in parentheses; p-values in square brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.