STRATEGIC CONFIDENCE? THE GENDER CONFIDENCE GAP AND ANTICIPATED DISCRIMINATION IN ECONOMICS PEER-REVIEW*

Margaret Samahita† Martina Zanella‡

This paper examines gender differences in stated confidence in a novel and consequential setting – the peer review of academic conference submissions. Contrary to expectations, female reviewers report higher confidence than males. Results suggest that stated confidence is context-dependent and may be strategically used by women in response to anticipated bias. This behavior reflects a rational response to the undervaluation of their expertise: organizers' decisions tend to align more closely with male reviewers' recommendations, even if female reviewers' evaluations are more predictive of actual paper quality.

*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. The data can be obtained by filing a request directly with the Irish Economics Association. Contact information can be found at `https://www.iea.ie/contact-the-iea/`. The project received IRB approval from the UCD Human Research Ethics Committee in February 2024 (reference number HS-LR-24-103-Samahita). 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 examine stated confidence in a novel and consequential setting: the peer review of academic conference submissions. Specifically, we study whether gender differences exist in the confidence reviewers declare in their expertise when evaluating a paper. This setting is interesting for three reasons. First, expressing confidence when evaluating others – rather than oneself – may involve lower reputational and identity costs for women, for whom appearing assertive may go against societal norms. Second, stated confidence in this context carries higher stakes as it may affect the person whose work is being evaluated. Third, because reviewer assessments inform organizers' acceptance decisions, and perceptions of reviewer credibility may vary by gender, this is a setting in which male and female reviewers may differ in how confidently they choose to present their expertise.

We study organisers' and reviewers' decisions in the Irish Economic Association (IEA) Conference from 2017-2023.[1] Our empirical strategy exploits a distinctive feature of the review process. For each submission, the referee is asked not only to evaluate the paper across various dimensions but also to rate their confidence in their own expertise. To our knowledge, no other economics conference systematically collects this information, offering a rare opportunity to study stated confidence in a professional evaluation setting—one that may involve different incentives from those examined in previous research, as described above. We supplement these data with information on referees, authors, and papers, collected from CVs, institutional/personal webpages, and Google Scholar.

Our central—and surprising—finding is that female reviewers report *higher* confidence than male reviewers, even when reviewing similar papers and controlling for individual reviewer characteristics. What can account for this positive gender gap in confidence? Selection into the profession is the first explanation we explore. If, as some have argued, women must be more like men to

[1] The 2020 conference was excluded as it was cancelled due to the Covid-19 pandemic.

break the glass ceiling in male-dominated fields, we might expect gender differences in confidence to disappear, and even reverse at higher levels of seniority (Adams and Funk, 2012). Indeed, the positive gender confidence gap is especially pronounced among more senior women. However, our result cannot be fully explained by selection into seniority. When we test this selection hypothesis for two characteristics likely correlated with confidence - namely expertise and experience, we find that senior women do not declare significantly higher confidence when evaluating papers within their own field of expertise (only for papers outside their field of expertise), and the positive gender gap persists even among referees with zero reviewing experience. These findings are inconsistent with selection into seniority being the sole explanation for the observed patterns.

We propose that female reviewers may strategically state higher confidence because they anticipate their expertise is more likely to be questioned by the organizers. Several patterns in the data support this interpretation. First, as noted above, the gap is driven by papers outside the reviewers' field of expertise, where perceived credibility may be lower. Second, the gap is larger when reviewers are less familiar with the organizers—a context that increases information asymmetry and may trigger statistical discrimination (e.g. Phelps, 1972; Ashenfelter and Albert Rees, 1973; Reuben et al., 2014; Bohren et al., 2019; Coffman et al., 2021). Third, in line with the statistical discrimination rationale, both female and male reviewers declare higher confidence when reviewing papers in stereotypically incongruent fields. Fourth, female reviewers show especially high confidence when evaluating male-authored papers, which may trigger identity-related concerns or stereotype threat.

Finally, we find suggestive evidence that female reviewers' higher stated confidence may reflect a rational response to the actual devaluation of their expertise. We find that, controlling for paper and reviewer characteristics, organizers' decisions more closely align with male referees than female referees for marginal papers. Notably, this gender influence gap persists even among confident or senior reviewers. These findings suggest that women's expertise is not only undervalued, but also that their strategic expression of confidence may be a justified and adaptive response.

To assess whether organisers' bias leads to suboptimal decisions, we examine how reviewers' evaluations correlate with the paper's subsequent success. We compare the number of citations accumulated in the years following the conference for papers evaluated by reviewers of different genders that obtained the same evaluation score, net of reviewer and paper characteristics. Female reviewers' evaluation of the paper appears to correlate more positively with the paper's long-term citations ceteris paribus. In other words, female reviewers' evaluations are, if anything, more predictive of paper quality—despite being less influential.[2]

Our paper contributes to three strands of literature. First, the paper enriches our understanding of the existence, magnitude, and drivers of the gender gap in confidence. Many papers document

[2]Consistent with this, we find no evidence that female reviewers are more lenient than males in their evaluations.

that women exhibit lower levels of confidence (Jakobsson, 2012; Barber and Odean, 2001; Buser et al., 2014), especially in male 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] The gender gap in confidence is also observed when subjects are asked to evaluate an issue or offer an opinion (for example, Gordon and Dahl, 2013; Sarsons and Xu, 2021; Sievertsen and Smith, 2025b; Coffman, 2014). Our paper challenges the narrative that women are uniformly underconfident by documenting that gender gaps in stated confidence are highly context-dependent. Unlike previous studies focused on self-assessments or personal opinions, we examine confidence while evaluating *others'* work—a setting where identity costs likely differ and incentives may exist to challenge the societal expectation prescribing women to be less assertive.

Our paper also adds to the literature on strategic responses to discrimination. Many studies show that women's expertise is often undervalued relative to men's: in the economics profession (e.g. Doleac et al., 2021; Hospido and Sanz, 2021; Dupas et al., 2021, Sarsons et al., 2021, Koffi, 2021, Mengel et al., 2019; Boring, 2017; Boring and Philippe, 2021, Lundberg and Stearns, 2019; Baltrunaite et al., 2022; Eberhardt et al., 2023, Card et al., 2020; Hengel and Moon, 2020; Hengel, 2022) and in other domains (e.g. medicine: de Vaan and Stuart, 2022; Sarsons, 2017; finance: Klein et al., 2021; hiring: Radbruch and Schiprowski, 2023). Yet few studies explore how disadvantaged groups strategically respond to anticipated bias, particularly in real-world, professional settings (He et al., 2024; Kudashvili and Lergetporer, 2022; Biavaschi et al., 2017; Kang et al., 2016; Zussman, 2013; Arai and Skogman Thoursie, 2009). The results of this paper provide insights into one such response strategy not previously documented—confidence can be strategically manipulated to respond to anticipated discriminatory treatment.

Third, we contribute to research on the economics profession itself. Many studies have documented gender disparities in the economics profession, cited in the previous paragraph. Our paper documents the gender gap in influence in a novel setting, that of referees' recommendation to conference organizers. Moreover, we examine how women navigate and respond to these gender gaps, showing that gendered expectations shape how expertise is signaled and interpreted.

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 evidence on the gender confidence gap. In Sections IV and V we investigate the explanations behind the observed patterns. Section VI discusses the evidence on the gender influence gap, while Section VII discusses the implications for paper outcomes. Finally, Section VIII concludes.

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

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 5). Submission closes in February, after which the conference organizers assign papers to a pool of around fifty reviewers according to their field of research. Reviewers include economists from Irish and non-Irish institutions as well as the conference organizers, who review papers when no reviewer can be found (for example when the assigned referee does not submit their review in time).[4] Around 60% of submitted papers are accepted every year.

The revision process Each paper is assigned to a reviewer. Each reviewer typically evaluates three papers, though conference organizers often review more. 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 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 measures 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." Screenshots of the webpage that reviewers are presented with can be found in Appendix Figures A.1 and A.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.[5]

[4] We can only observe the final reviewer of the paper. We do not know whether the paper was initially assigned to a different reviewer who failed to submit the evaluation.

[5] 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.

The dataset We use data from the 2017–2023 conferences.[6] 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.[7] It includes details provided by the submitter, such as the paper's title, abstract, primary field, and optional secondary field, 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).[8]

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 1 and 2, presented at the reviewer-year level to account for referees 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. 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.

[6]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.

[7]2021 organizers manually deleted papers that were withdrawn by the authors after acceptance and we were not able to fully recover the identity of the reviewers, resulting in 2 missing observations.

[8]47 papers have missing reviews. These papers were evaluated directly by the organizers when referees failed to submit their reviews by the deadline. The organizers made acceptance or rejection decisions without assigning an acceptance score.

Table 1- 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: Full 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, accounting for institutional changes and promotions. 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 2- 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 information about the referee (31, 1, 30 reviewer-year pairs, respectively). The results are robust to replicating the analysis excluding these observations.

Table 3 presents descriptive statistics on paper allocation across referees. On average, about 40% of reviewer-paper matches are within the same research field, with no significant difference across reviewer genders. Each paper has, on average, 72% male authors, and this proportion also does not differ significantly by reviewer gender. Similarly, the proportion of Irish authors is consistent across genders. However, male reviewers evaluate significantly more papers from authors with a PhD than female reviewers. Few referees review papers written by their colleagues, and very few papers had citations before the conference year.

Table 3- Paper Allocation

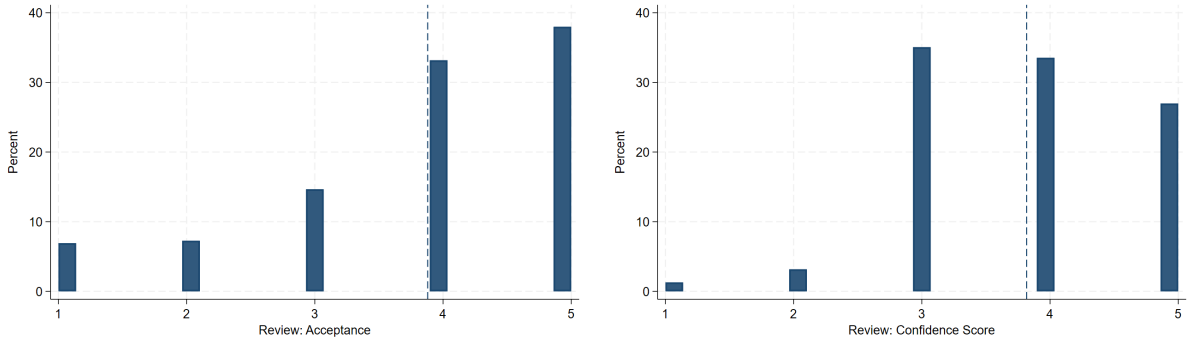
	Male Refs		Female Refs		T-test
	N.	Mean	N.	Mean	Diff. p-value
Within field of expertise	622	0.42	231	0.37	0.15
Within field of expertise (not organizers)	503	0.45	187	0.42	0.47
Male submitter	622	0.71	231	0.69	0.62
Prop. of male authors	622	0.72	231	0.72	0.88
Prop. of authors with PhD	622	0.74	231	0.66	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 referee'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.

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 reported high confidence overall, with an average score of 3.82 (SD = 0.91). Most reviewers (60%) were either confident (34%) or very confident (27%), while 35% were somewhat confident and 4.50% had low or no confidence.

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 The Gender Gap in Confidence

To identify any gender gap in confidence, we compare the confidence scores provided by reviewers of different genders controlling for an array of characteristics of the paper and the reviewer. We estimate:

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

where Y_{ipt} denotes the score reviewer i gave to paper p in year t. We control for a set of reviewer-specific controls: 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: 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 WP series, a dummy equal to one if submitter's institution is the same as the referee's, and N. citations of the paper before the conference.[9]

[9]The number of publications, citations, and years post-PhD are imputed to average if we could not find information about the referee (31, 1, 30 reviewer-year pairs, respectively). We include a dummy for missing observations. The

Unsurprisingly, a confidence gap exists. However, differently from what has been previously documented in the literature (Sarsons and Xu, 2021; Jakobsson, 2012; Barber and Odean, 2001; Lundeberg et al., 1994), female reviewers declare to be *more* confident than male reviewers regarding their knowledge in the topic discussed in the submission. On average, female reviewers' confidence is 3.99 (on a 1-5 scale) while the corresponding score is 3.75 for male reviewers (t-test, p=0.0009), as shown in Figure 14. The gap is robust to controlling for the characteristics of the paper and the referee – Table 4. Women on average declare 0.18 higher confidence scores (corresponding to 5% of the average confidence score 3.82, or 0.23 standard deviations) compared to men when evaluating papers with similar characteristics, holding the characteristics of the referee constant.

Table 4- 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). The number of publications, citations, and years post-PhD are imputed to average if we could not find information about the referee (31, 1, 30 reviewer-year pairs, respectively). Column (5) includes only the referee-paper pairs for which we have a complete set of information. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

IV The role of selection into the profession

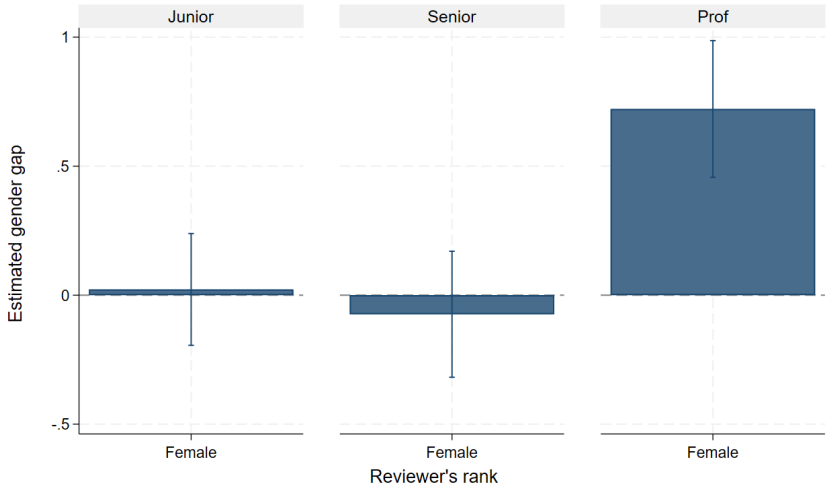
What can account for this surprising positive gender gap in confidence? Selection into the profession can be one explanation. If women must be more like men to succeed in male-dominated fields, we might expect gender differences in confidence to disappear in these settings. They might even reverse at higher levels of seniority (Adams and Funk, 2012), particularly if females who make it to the top had to overcome barriers to entry, resulting positively selected in terms of expertise, or other traits correlated with confidence. We therefore check for heterogeneity in the gender confidence gap by seniority in the profession.

The results obtained from estimating equation (1), including interactions of the referee's gender

main results are confirmed if we replicate the analysis by excluding these observations, as shown in Table 4. 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.

with their rank, confirm this expectation. Figure 2 shows that the average estimated difference in confidence is indeed driven by reviewers who are Professors or equivalent, while the positive gender effect is not present for junior economists. Table 6 provides the coefficients behind Figure 2.

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 referee 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 full Professor or equivalent. The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals.

Can then selection into seniority explain the observed gender confidence gap? While senior women can be selected along various traits, our data enables us to test this selection hypothesis for two such characteristics correlated with confidence - namely expertise and experience.

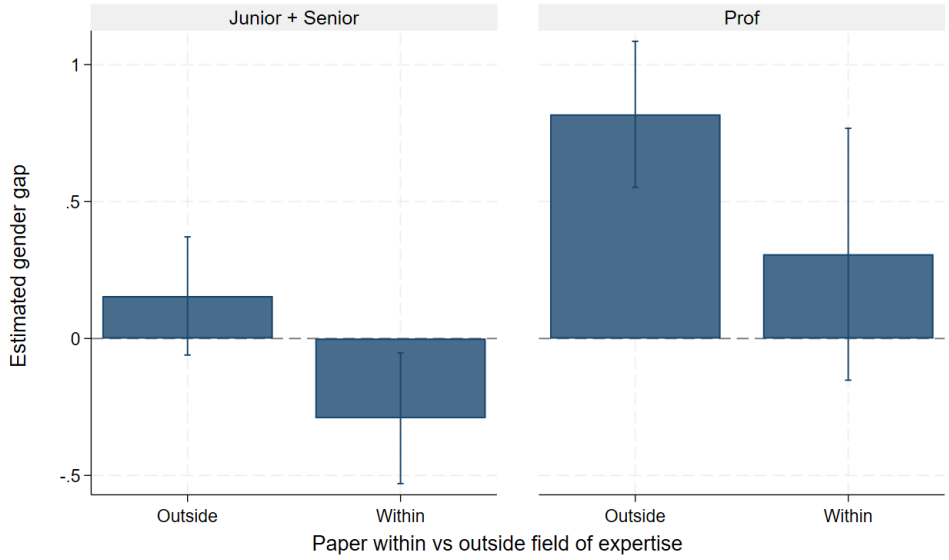
First, we test whether female Professors state higher confidence in their knowledge of the papers' topics because they are positively selected in terms of their expertise. We do this by estimating the gender confidence gap both for papers within and outside the field of expertise. We measure the field of expertise using a dummy equal to 1 if the paper's primary field, as declared by the author, matches one of the referee's fields of interest, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar.[10] We then estimate specification (1) including the interaction of this expertise dummy with the female dummy and the professor dummy to estimate the effect of expertise on confidence for referees of different genders and rank.

Figure 3 shows that female professors do not display significantly higher confidence compared

[10]Table 3 shows that the share of papers within and outside the field of expertise allocated to reviewers does not systematically vary by gender. In Appendix B we provide a different measure of field expertise using LLM, our results are robust to this alternative definition.

to male professors when the paper is within the field of expertise.[11] Indeed, the positive gender gap is driven only by papers outside the field of expertise. If women were positively selected in terms of expertise and this was the sole explanation for their higher confidence, we would expect the positive confidence gender gap to be present also for papers within the field of expertise. Note that, when the paper is within the field of expertise, a negative gender confidence gap appears for more junior reviewers. We will come back to this in the next section.

Figure 3- Selection: The role of expertise



Notes: Bar graph that displays the estimated gender confidence gap (female minus male) for papers outside and within the field of expertise by rank, net of paper and referee 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 referee's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. 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 full Professor or equivalent. The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals.

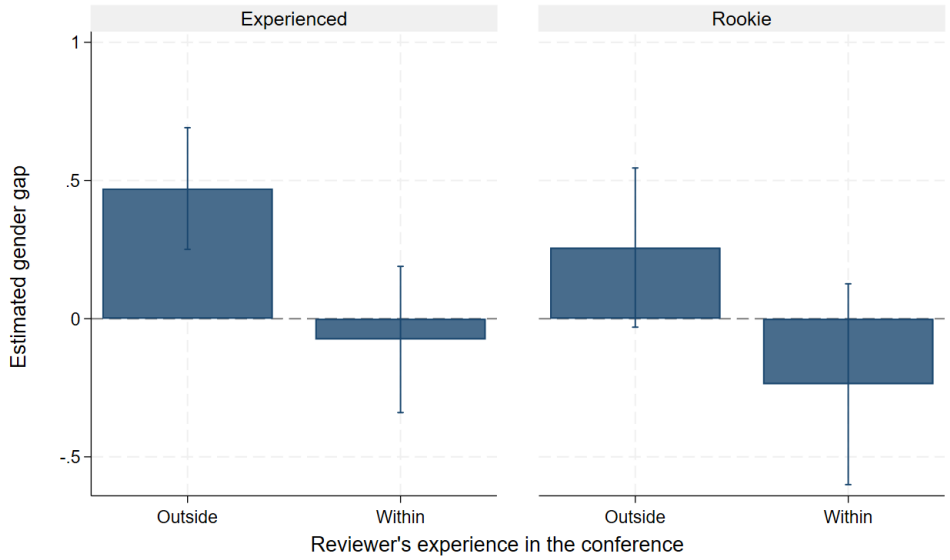
Second, we examine whether the positive gender confidence gap can be explained by senior women having greater experience in reviewing papers. We use as proxy the number of papers reviewed in previous conferences, which appears to be strongly correlated with declared confidence (right panel of Figure 15). We find that the rank in the profession is indeed positively correlated with reviewing experience more so for women than for men (left panel of Figure 15).[12] However, when we estimate the gender confidence gap for reviewers with different levels of reviewing ex-

[11] Table 7 provides the results behind the figure.

[12] 1 is Ph.D. student or equivalent, 2 is Post-doc (or Research Fellow) or equivalent, 3 is Assistant Professor or equivalent, 4 is Associate Professor or equivalent, and 5 is Full Professor or equivalent.

perience, we find that even rookie referees—those with no prior reviewing experience—exhibit a positive gender confidence gap when evaluating papers outside the field of expertise (Figure 4 and Table 8). This suggests that while reviewing experience helps to explain females' higher confidence, it does not fully account for the observed positive gender gap in confidence.

Figure 4- Selection: The role of experience in reviewing



Notes: Bar graph that displays the estimated gender confidence gap (female minus male) for papers outside and within the field of expertise by experience in reviewing, net of paper and referee 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 referee's fields, as stated in personal and institutional webpages, CVs, LinkedIn profiles, and Google Scholar. Rookie is a dummy equal to one if the reviewer has never reviewed a submission for the conference before. The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals.

V Strategic motive

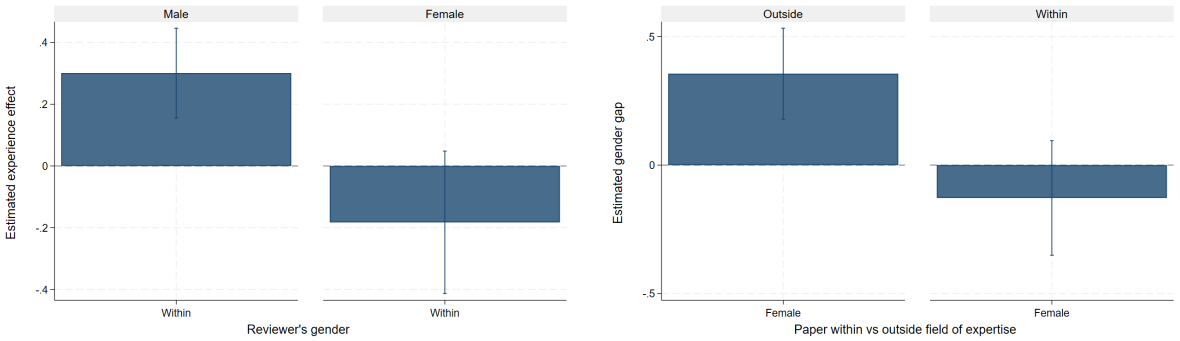
In this section, we argue that female reviewers may strategically state higher confidence, anticipating that organizers would be more likely to doubt their expertise compared to male reviewers'.

We proceed by providing four pieces of evidence that all together support the hypothesis of strategic motives. The gender confidence gap is stronger in cases when women anticipate their expertise to be questioned: (i) when reviewing papers outside their field of expertise, (ii) when they are less familiar with the organizers, (iii) when reviewing papers in stereotypically male fields, and (iv) when reviewing male-authored papers, a situation which may trigger identity considerations and stereotype threat.

Field of expertise For the first piece of evidence, we come back to the distinction between papers within and outside the field of expertise introduced in section IV. While reviewers may have a good general knowledge of economics and regularly review papers outside their specific field of expertise, we would expect confidence to be higher when reviewing papers in their own fields of expertise if no strategic considerations exist.

This prediction is met for male but not female referees. We estimate specification (1) including the interaction of this expertise dummy with the female dummy to estimate the effect of expertise on confidence for referees of different genders. Figure 5 displays the results of this exercise. Men declare higher confidence when the paper is within the field of expertise, but this is not the case for women (panel a). As a consequence, the gender confidence gap is driven by papers outside the referees' fields of expertise (panel b). Indeed, female referees declare significantly higher confidence compared to men only when the paper is outside their field of expertise. Table 9 provides the complete set of estimates Figure 5 refers to.

Figure 5- The role of expertise



(a) Estimated experience effect **(b)** Estimated gender gap

Notes: Bar graph that displays the estimated effect of being an expert on the paper topic 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 referee 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 referee'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.

This result is again surprising. Other papers studying the Economics profession find that senior female economists are less confident than men when responding to questions or providing opinions about the economy and government policy outside their field of expertise (Sarsons and Xu, 2021; Sievertsen and Smith, 2025b). These papers, however, study contexts that are public and are hence characterized by higher reputational and identity costs: females may not want to appear overconfident for fear of backlash (Rudman and Phelan, 2008). In our (more private) setting, the reputational

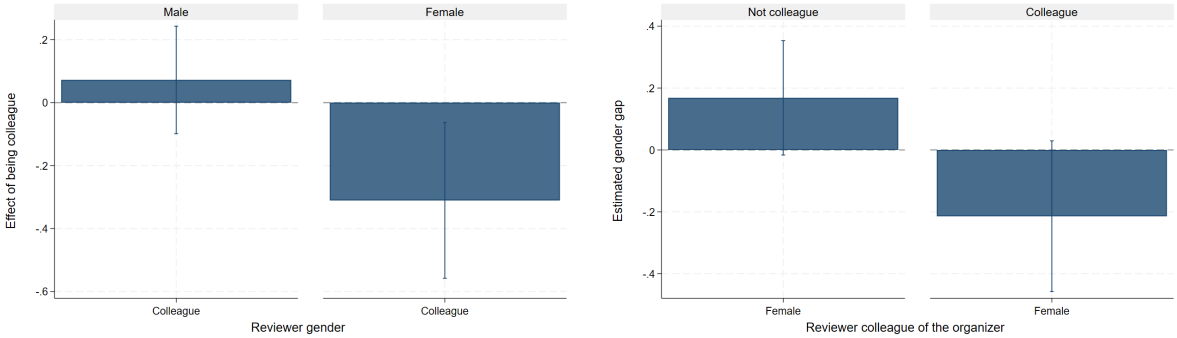
cost of appearing overconfident is likely exceeded by the potential benefits that can derive from declaring higher confidence, particularly when the reviewed paper is outside the field of expertise and women may anticipate their expertise to be discounted. In line with this reasoning, as we saw in Figure 3, the traditional negative gender gap (male reviewers stating higher confidence) appears when junior referees review papers within their field of expertise. For them, the reputational costs might be higher relative to more senior referees who can afford to state higher confidence, even in private settings. Additionally, the benefits of overstating confidence may be lower when there is less anticipated bias – i.e. when reviewing papers within the field of expertise.

Colleagues of the organizers An alternative situation where female reviewers might perceive a stronger need to overstate their confidence is when they are not familiar with the conference organizers. Indeed, they might anticipate that lack of information or uncertainty regarding their expertise might trigger organizers to statistically discriminate, exacerbating the reliance on priors and stereotypes when forming expectations about competence (e.g. Phelps, 1972; Ashenfelter and Albert Rees, 1973; Reuben et al., 2014; Bohren et al., 2019; Coffman et al., 2021).

Panel (a) of Figure 6 confirms this prediction. Female referees' declared confidence depends on their familiarity with the organizers. The same is not true for men. We compare the confidence score declared by female and male reviewers, distinguishing between those who are colleagues (defined as working in the same institution) and non-colleagues of the organizers, net of paper and reviewers' characteristics as described in equation (1). We exclude reviewers who are the conference organizers themselves. Note that the organizing committee changes every conference, hence we are not comparing the same set of individuals every year. We find that women state higher confidence when they are not colleagues of the organizers, compared to when they work in the same institution as the organizing team. On the other hand, being in the same institution as the organizers does not make a difference for men (panel a).

Hence, the observed pattern of positive gender confidence gap (female reviewers declaring higher confidence compared to males) is driven by reviewers who work in a different institution from the organizers (Figure 6, panel b). Table 10 provides a formal test and additionally shows that when reviewers are colleagues with the organizers and review papers within their field of expertise, the traditional negative gender gap (male reviewers stating higher confidence) appears. This provides additional evidence that the higher stated confidence of women is a strategy rather than a trait that distinguishes the women in our study from those studied in other articles. Indeed, in this "safer" context, women might perceive there is neither need nor scope to strategically overstate their confidence as their recommendations are less likely to be ignored by colleagues familiar with their expertise.

Figure 6- The role of familiarity to organizers



(a) Estimated effect of being colleague **(b)** Estimated gender gap

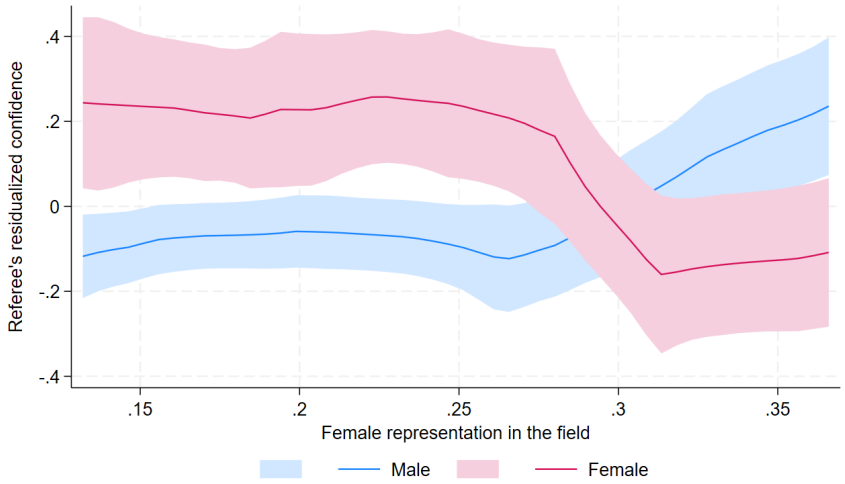
Notes: Bar graph that displays the estimated effect of being a colleague of the organizers for men and women (panel a) and the estimated gender confidence gap (female minus male) for non-colleagues and colleagues of organisers (panel b), net of paper and referee controls as specified in equation (1). The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals.

Stereotypically male fields Closely related to the previous mechanism, if women are responding strategically to anticipated statistical discrimination, we can expect them to overstate their confidence more when the paper belongs to a stereotypically male topic. To test this prediction, we categorize papers based on the "stereotypicality" of the paper's primary field (as indicated by the authors), defined as the share of women working in the field.[13] 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 11.

Figure 7 clearly shows that this prediction is empirically met. The figure displays the correlation between the share of women working in the paper's primary field and the confidence score provided for each paper by male and female reviewers, net of paper and reviewers' characteristics, as described in equation (1). It can be clearly seen that women declare higher confidence ceteris paribus when the paper's topic belongs to a gender-incongruent field. The same pattern is present for men, even though it is stronger for women. The patterns are formally tested in Table 12.

[13]The relationship between stereotypes and gender segregation of the workforce has been documented in psychology (e.g. Garg et al., 2018), and the share of females as an indicator of friendliness/stereotypes of the sector/field has been widely used in the literature (e.g. Zanella, 2024; Bostwick and Weinberg, 2022; Kugler et al., 2021; Hebert, 2020).

Figure 7- The role of gender stereotype of paper topic



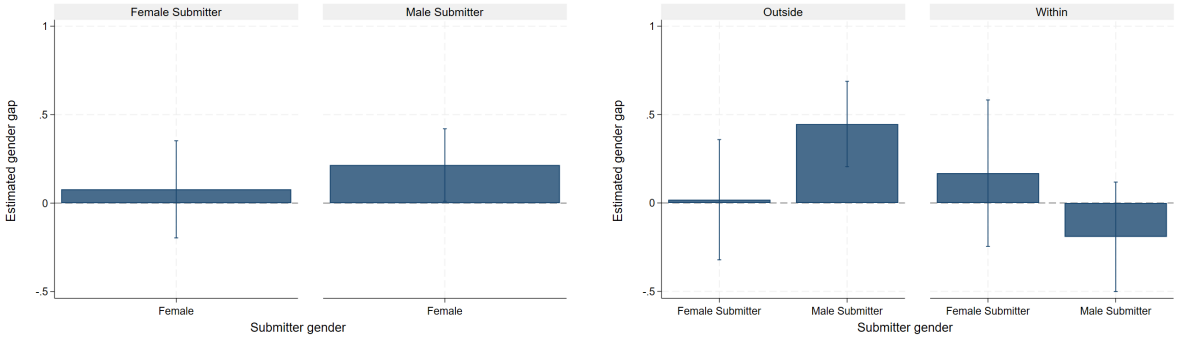
Notes: The graph displays the correlation between the share of women working in the paper's primary field (x axis) and the confidence score provided by male (blue line) and female (red line) reviewers (y axis) for each paper. The red and blue lines display the smoothed values of a kernel local polynomial regression of the residuals from a regression of the reviewers' confidence score on paper and referee characteristics, as described in equation (1). 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 11.

Authors' gender Finally, female reviewers may strategically declare higher confidence when certain characteristics of a paper make their gender salient, triggering identity-related considerations—for example, when reviewing papers authored by men.

This is indeed what we find. Figure 8 shows the gender confidence gap by the gender of the submitter, for unblinded submissions, net of paper and reviewers' characteristics as described in equation (1). Indeed, a stronger gender confidence gap appears when the submitter is male (panel a).[14] In particular, this gender gap is again driven by papers outside the reviewers' field of expertise, suggesting that it might not be enough to be reminded about gender identity to trigger a strategic response. This becomes relevant when female reviewers perceive additional reasons for their recommendations to be undervalued. Table 13 presents a formal test, including also an alternative definition of male-authored papers, based on the proportion of male authors. The results are qualitatively unchanged.

[14]Table 3 shows that the share of papers with male submitters allocated to reviewers does not systematically vary by gender.

Figure 8- The role of the submitter's gender



(a) Estimated average gender gap **(b)** Estimated gender gap by field of expertise

Notes: Bar graphs that display the estimated gender confidence gap (female minus male) by paper submitter's gender (panel a) and separated by field of expertise (panel b), net of paper and referees' controls as specified in equation (1). The figure is generated using Stata's *marginsplot* command. Bars indicate 95% confidence intervals.

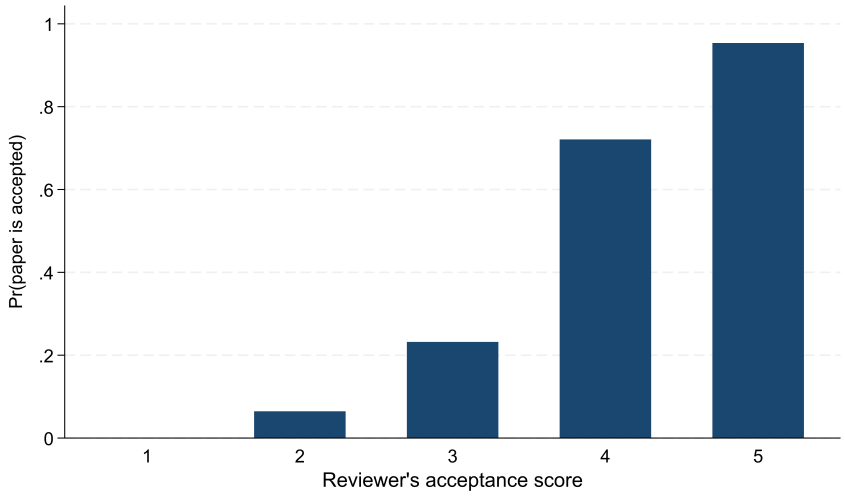
We have thus far presented four scenarios in which one should expect, and indeed we find, that the motive for overstating confidence by female reviewers is stronger. While no single piece of evidence is conclusive, these multiple data points taken together provide support for the explanation that the gender confidence gap is driven by a strategic motive: female reviewers *anticipate* that their expertise may be undervalued. In the next section, we will check whether or not the expectation of discrimination is indeed correct.

VI The Gender Influence Gap

The analysis on confidence suggests that female reviewers strategically overstate their confidence in anticipation of their expertise being discounted. Does this reflect a rational response to an *actual* devaluation of their expertise? To provide an answer to this question, we estimate whether there exists a gender gap in referee influence on organizer decisions by studying the relationship between the acceptance score assigned by reviewers and the organizers' decision about the paper.

The acceptance score assigned by reviewers significantly influences organizers' decisions, as shown in Figure 9. Papers with a score of 2 or lower ("do not accept" or "weak reject") have low acceptance rates: 0% and 6%, respectively. Among papers with a score of 3 ("borderline accept/reject"), 23% are accepted. Papers receiving a score of 4 ("probably accept") have a 72% acceptance rate, while those with a score of 5 are highly likely to be accepted, with a success rate of 95%.

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



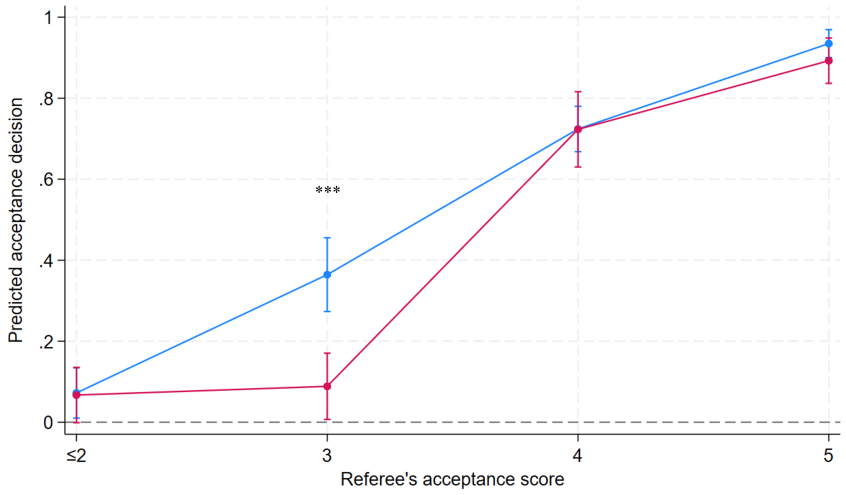
Notes: Share of papers accepted to the conference (y axis) by reviewers' acceptance scores (x-axis).

However, papers that receive similar scores from male and female reviewers do not have the same probability of being accepted. To identify whether a gender gap in referees' influence on organisers' decisions exists, we compare the acceptance decision by organisers for papers refereed by male versus female referees, controlling for an array of characteristics of the paper and the reviewer.

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

where Y_{ipt} denotes the probability that organizers in year t accept paper p for which reviewer i provided acceptance score j. We control for the same set of reviewer-specific and paper-specific characteristics as described for equation (1). Figure 10 plots the predicted acceptance probability for each referee's acceptance score, net of paper and referee characteristics, resulting from this exercise. While there is no gender gap in papers' likelihood of acceptance when referees recommend rejecting (acceptance score equal to 1 or 2) or accepting (acceptance score equal to 4 or 5), a significant gender influence gap exists for papers for which reviewers are undecided (score equal to 3). In particular, the probability that the paper is accepted by the organizers when the reviewer is undecided is higher if the reviewer is a man. A formal test confirming the results can be found in Table 14.

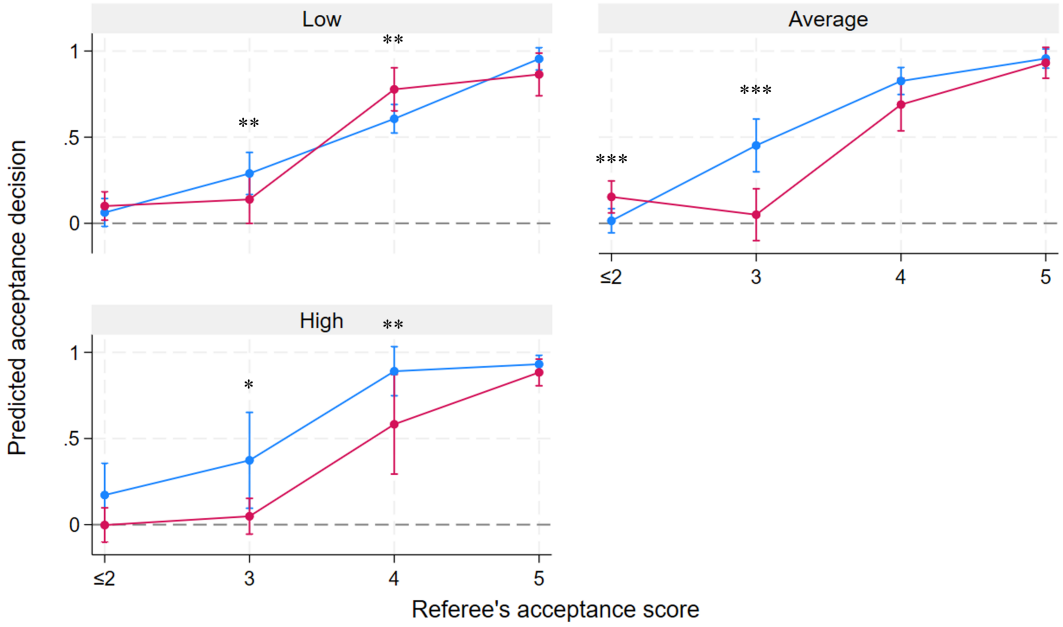
Figure 10- Gender gap in the probability of the paper being accepted



Notes: Average predicted probability that the paper is accepted by the organizers, plotted against reviewer's acceptance decisions, net of referee and paper characteristics as in equation (2). The figure is generated using Stata's *marginsplot* command. Stars indicate the significance of the gender gap for each acceptance score using robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We interpret this result as male reviewers having greater influence than female reviewers. The reason is threefold. First, if organizers perfectly followed the reviewers' decisions, they would accept papers with 50% probability when reviewers are undecided (a random draw). The organisers' likelihood of acceptance differs from random by 18 percentage points for male referees and 46 percentage points for female referees. Secondly, the pattern is stronger the higher the confidence declared by the reviewers, as seen in Figure 11 (estimates provided in Table 15). Notably, a negative gender gap appears also for papers for which reviewers suggest probably accepting (acceptance score equal to 4) for average and high (significant) levels of declared confidence. Finally, this cannot be reconciled with the reviewers' belief that women are more lenient in their evaluations compared to men. Indeed, Table 18 provides evidence that women do not provide higher scores than men. Although we cannot completely rule out the possibility that organizers hold incorrect beliefs regarding gender differences in the leniency of evaluations, it is unclear why these beliefs would increase with reviewers' confidence—a necessary condition for this explanation to fully reconcile the observed patterns.

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



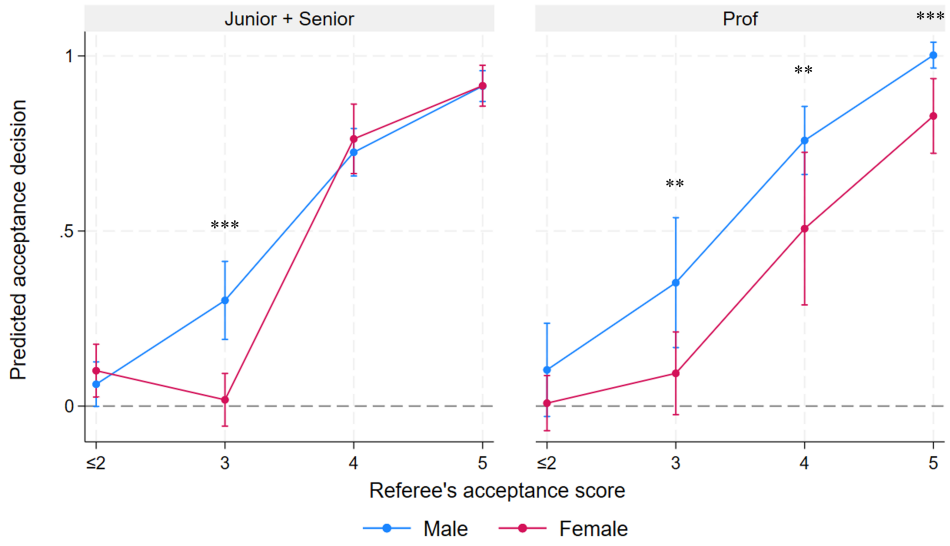
Notes: Average predicted probability that the paper is accepted by the organizers, plotted against reviewer's acceptance decisions, net of referee and paper characteristics as in equation (2). Each graph displays results for a subsample of referees with different confidence: reviewers whose confidence 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. Stars indicate the significance of the gender gap for each acceptance score using robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Lastly, the gender influence gap is particularly strong for Professors. As shown in Figure 12, female professors' recommendations are less likely to be followed than male professors across all acceptance scores.[15] This finding is particularly interesting since the literature suggests that the gender gap in the valuation of expertise may be mitigated by credentials as they may help signal women's expertise to evaluators (Mengel et al., 2019; Ayalew et al., 2021; Bohren et al., 2019; Sievertsen and Smith, 2025a).[16] This is clearly not happening in this case. Hence, a strategic expression of confidence appears to be a potentially justified and adaptive response. Indeed, Professors are the subsample of referees that declare the highest confidence.

[15]The gender gap becomes less precisely estimated when we restrict the sample to Professors as the number of observations falls to 254.

[16]For example, the gender gap in teaching evaluation is reduced for more senior professors (Mengel et al., 2019). The signal may even be stronger as, if women face discrimination, successful women may be perceived as extraordinarily good (Bohren et al., 2019). Sievertsen and Smith (2025a), for example, show that individual expert opinions are more persuasive when they are expressed by senior female economists. While this is contrary to expectation, the authors attribute the finding to the saliency of the women's credentials as elite economists.

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 referee and paper characteristics as in equation (2). Each graph displays results for a subsample of referees: up to and including Associate Professors or equivalent (left) and Professors or equivalent (right). The figure is generated using Stata's *marginsplot* command. Stars indicate the significance of the gender gap for each acceptance score using robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

VII Is this consequential? Paper outcomes

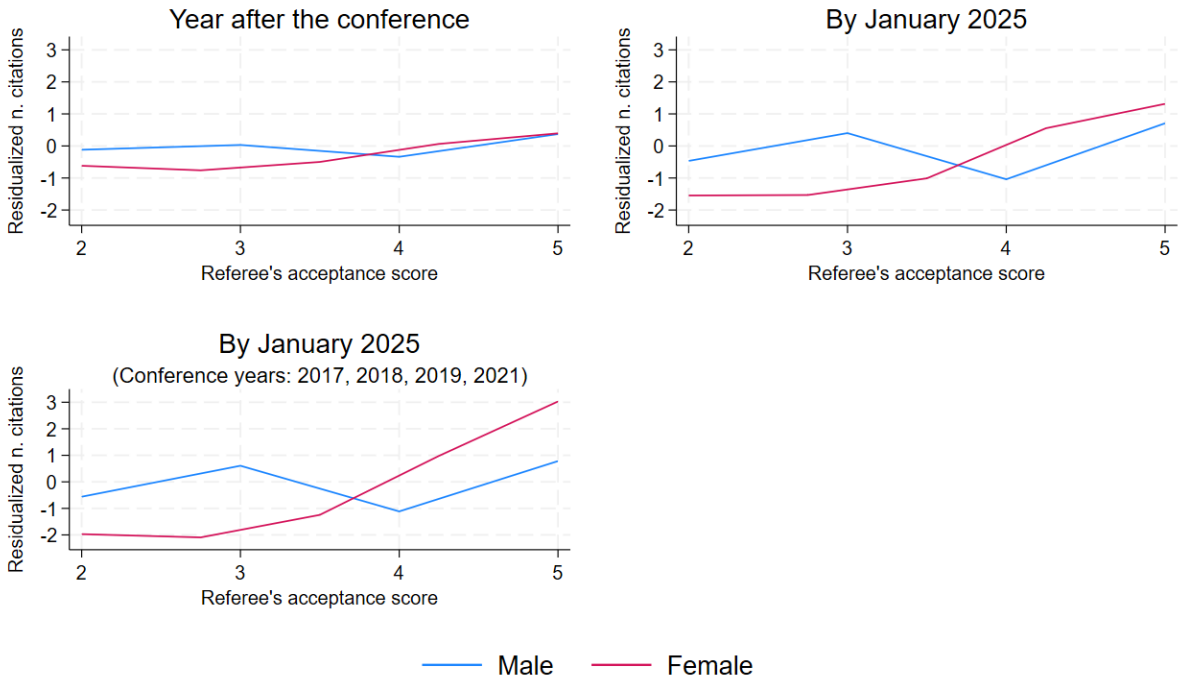
Taken together, the findings indicate that, despite declaring significantly higher confidence, female reviewers' evaluations are followed less closely by the organizers. What remains to ask is: is this behavior optimal from the organizers' point of view? To answer this question, we plot the correlation between the paper's success, proxied by its post-conference citations, and the reviewer's acceptance scores in Figure 13. We do this considering (i) the number of new citations accumulated by the paper the year after the conference, (ii) the number of citations accumulated by the paper by January 2025, and (iii) the number of citations accumulated by the paper by January 2025 only for papers submitted to conferences prior to 2022. The latter are papers for which we can infer more long-term dynamics as at least 3 years have passed since the conference.

Female reviewers' evaluations appear to correlate more positively with paper success. Notably, the graphs show an overall steeper slope for female reviewers compared to males. For low-acceptance-score papers, female-refereed papers accumulate a lower number of citations. The gender gap reverses for high-acceptance-score papers, where female-refereed papers accumulate a higher number of citations. The gender gap in correlation is negligible the year following the

conference, but becomes more and more evident the further away in time we move from the conference.

While these patterns are only suggestive, they point towards female referees providing more accurate evaluations of the papers. Indeed, the negative gender gap in citations among low-acceptance-score papers—particularly those receiving a score of 3—can be partly attributed to organizers' higher likelihood of accepting papers reviewed by male referees, thereby increasing their exposure and subsequent citations. However, this citation gap widens over time. As the influence of the conference diminishes, citation counts are likely to more accurately reflect the intrinsic quality of the paper. Similarly, the positive gender gap observed among high-acceptance-score papers also grows over time. Taken together, these patterns suggest that female referees' evaluations are more strongly correlated with the actual quality of the paper, pointing to female referees' expertise being sub-optimally undervalued.

Figure 13- Citations post conference



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

VIII Conclusion

In this paper, we document the existence of a gender gap in confidence and influence in a novel and consequential setting: peer review of academic conference submissions. Using data from the Irish Economic Association (IEA) Conference from 2017-2023, we find that female referees report *higher* confidence than their male counterparts—particularly when evaluating papers outside their field, when they are less familiar to organisers, and when reviewing papers in stereotypically male fields or authored by men. These patterns suggest that women may strategically overstate their confidence in anticipation of having their expertise discounted.

We find suggestive evidence that female reviewers' higher stated confidence may reflect a rational response to the actual devaluation of their expertise. Female reviewers have less influence than male reviewers: organisers' acceptance rates are more closely aligned with male referees' recommendations. This gender influence gap persists even after accounting for reviewer seniority and declared confidence.

These findings point to an inefficient and potentially self-reinforcing equilibrium. Referees' strategic overstatement of confidence may be an optimal individual response in a static setting, but could trigger a vicious cycle in which women's credibility continues to be questioned, and their stated confidence discounted. However, an analysis of papers' long-term outcomes suggests that organizers' behavior is suboptimal. Female reviewers' evaluations more closely correlate with the papers' accumulated citations 3 or more years after the conference, highlighting the importance of valuing the expertise of female referees.

Our paper focuses on a specific context—Irish economists and conference acceptance decisions. Nonetheless, the patterns documented here likely extend to other settings where individuals from underrepresented groups are tasked with evaluating others' work under conditions of potential bias. These include hiring and promotion decisions in the workplace, assessing candidates for loan applications, or providing a reference for a potential tenant. In such environments, strategic behavior such as overstating confidence can emerge as a rational response to anticipated discrimination. Moreover, while the Irish economics community is distinct, it is not anomalous. Women represent approximately 33% of economists in Ireland——a figure broadly consistent with gender representation in economics across Europe[17] and in other influential spheres such as politics and corporate governance.[18] In addition, the relatively small size of the Irish economics community

[17]This figure is roughly in line with the representation of women in economics in other European countries such as France and Italy (32% each), and the overall proportion for Europe at 32.5% reported in Auriol et al. (2022), though somewhat higher than the UK (26%) or the US (23%), as reported by IDEAS.

[18]The proportion of seats held by women in national parliaments in OECD countries was 33% in 2024 (World Bank statistics - link). The last European Commission Gender Balance on Corporate Boards Directive sets a target for EU large listed companies of 33% of the underrepresented sex among all directors - link.

likely fosters close professional networks and high interpersonal familiarity, which may attenuate the dynamics identified in this study.[19] In more anonymous or impersonal settings——where individuals are less known to one another——the gender influence gap may be more pronounced, and the strategic need to overstate confidence correspondingly greater.

References

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

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.

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

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.

Ayalew, S., S. Manian, and K. Sheth (2021). Discrimination from below: Experimental evidence from ethiopia. *Journal of Development Economics 151*, 102653.

Baltrunaite, A., A. Casarico, and L. Rizzica (2022). Women in economics: The role of gendered references at entry in the profession. Centre for Economic Policy Research Discussion Paper No. 17474.

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. *The Quarterly Journal of Economics 116*(1), 261–292.

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. *The Quarterly Journal of Economics 129*(3), 1409–1447.

[19]The *absolute* number of economists in Ireland is 345, much smaller compared to the above-named countries (France at 3,848, Italy at 3,517, UK at 3,954, and the US at 12,611) as reported by IDEAS.

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

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. *The 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.

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. S. Modestino, M. Niederle, J. Wolfers, et al. (2021). Gender and the dynamics of economics seminars. National Bureau of Economic Research Working Paper No. 28494.

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. and J. B. Kessler (2022). The gender gap in self-promotion. *The 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.

Garg, N., L. Schiebinger, D. Jurafsky, and J. Zou (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences 115*(16), E3635–E3644.

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. (2020). Gender stereotypes and entrepreneur financing.

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.

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.

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.

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.

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.

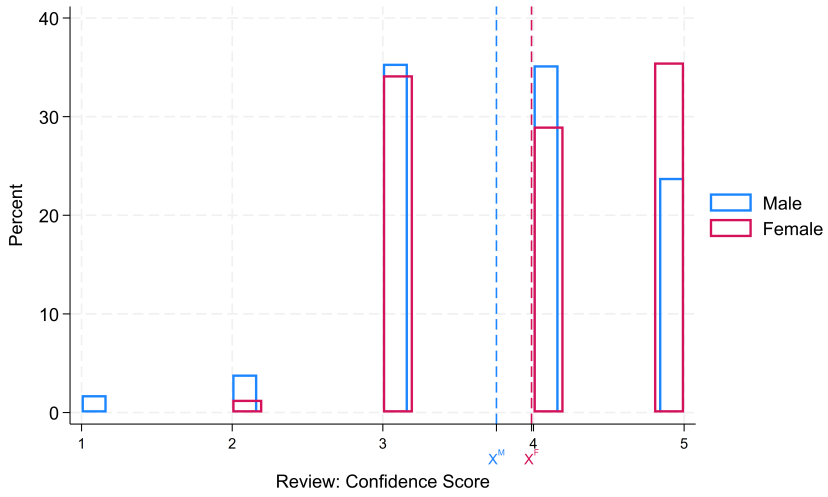
IX Additional Tables and Figures

Table 5- 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.

Figure 14- 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.

Table 6- Gender gap in confidence: the role of rank

	(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	0.51***	0.59***	0.80***	0.68***
	(0.16)	(0.17)	(0.17)	(0.18)
Senior	0.12	0.14	-0.05	-0.09
	(0.10)	(0.09)	(0.10)	(0.11)
Prof	0.36***	0.35***	0.17	0.21*
	(0.10)	(0.10)	(0.11)	(0.11)
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	0.58	0.69	0.82	0.73
	[0.00]	[0.00]	[0.00]	[0.00]

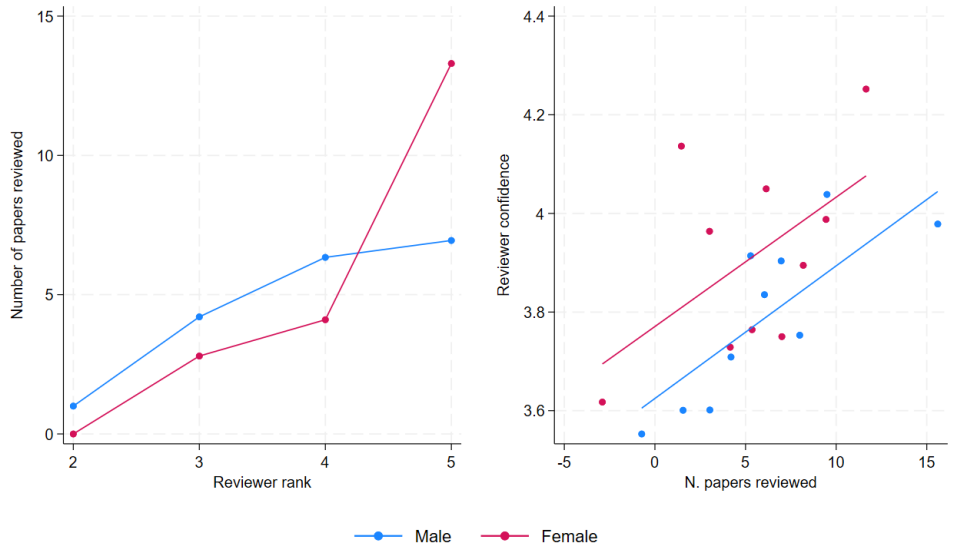
Note: This table reports the estimated gender gap in confidence by reviewer's rank. The regression specification is given by equation (1), interacting the female dummy with a Professor and Senior dummies. Senior includes rank 4, Professor is defined as rank 5. The omitted category is Junior - rank ≤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

Table 7- Gender gap in confidence: the role of expertise by rank

	(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]

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 referee'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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 15- 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).

Table 8- Gender gap in declared confidence: the role of experience

	(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

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 referee'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 9- Gender gap in confidence: the role of expertise

	(1)	(2)	(3)	(4)
Female	0.38***	0.42***	0.36***	0.41***
	(0.09)	(0.10)	(0.09)	(0.10)
Female x Within	-0.48***	-0.55***	-0.48***	-0.52***
	(0.14)	(0.15)	(0.14)	(0.15)
Within	0.33***	0.31***	0.30***	0.28***
	(0.07)	(0.08)	(0.07)	(0.08)
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 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 referee'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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10- Gender gap in confidence: the role of being familiar with the organisers

	(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
Year FE	X	X
Paper controls	X	X
Referee controls	X	X
Gender gap if Colleague=1	-0.21	
	[0.09]	
Colleague effect if Female=1	-0.31	
	[0.01]	
Gender gap if Colleague=1 & Within=0		-0.05
		[0.77]
Gender gap if Colleague=1 & Within=1		-0.47
		[0.02]

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 referee'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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11- 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.

Table 12- Gender gap in confidence: the role of the paper's field stereotypicality

	(1)
Female	1.07***
	(0.21)
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 11 in Appendix. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 13- Gender gap in confidence: authors' gender

	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		0.02		-0.20
		(0.14)		(0.18)
Within × Female		0.15		0.28
		(0.26)		(0.32)
Within × Male Authors		0.39***		0.68***
		(0.16)		(0.21)
Within × Male Authors × Female		-0.79***		-0.96***
		(0.31)		(0.39)
N	666	666	666	666
Year FE	X	X	X	X
Paper controls	X	X	X	X
Referee controls	X	X	X	X
Gender gap if Male Authors	0.22		0.18	
	[0.04]		[0.09]	
Gender gap if Within=0, Male Authors		0.45		0.43
		[0.00]		[0.00]
Gender gap if Within=1, Female Authors		0.17		0.31
		[0.42]		[0.21]
Gender gap if Within=1, Male Authors		-0.19		-0.25
		[0.23]		[0.12]

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 (columns 1), and the proportion of authors who are male (columns 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 (columns 2) and the proportion of authors who are male (columns 4), 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 referee'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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14- 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 information about the referee (31, 1, 30 reviewer-year pairs, respectively). Column (4) includes only the referee-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 15- Summary: The gender influence gap by declared confidence and rank

	(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 17 in Panels A-C and Table 16 in Panels D and E. P-values in square brackets.

Table 16- 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 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 18- Gender gap in reviewers' leniency

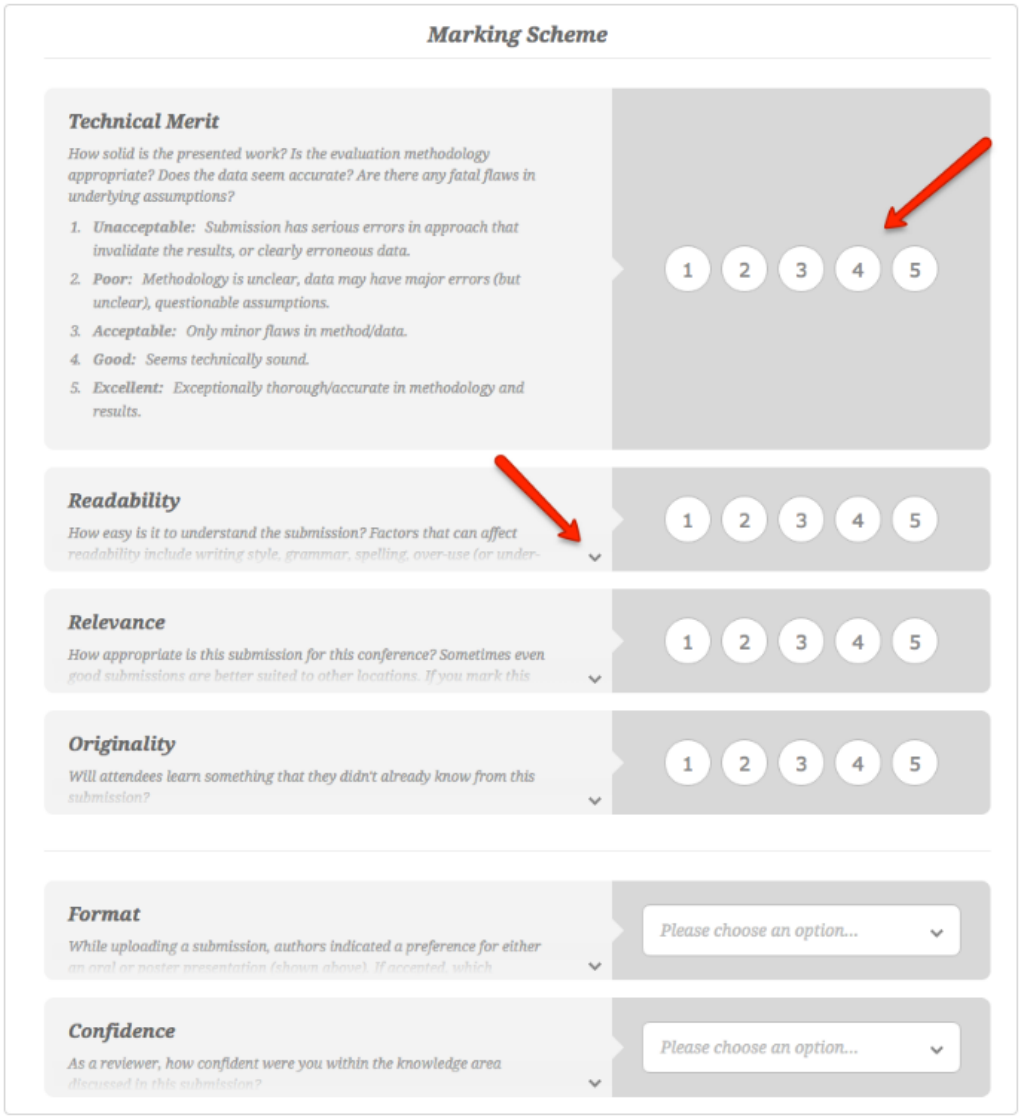
	Acceptance Score (1)	Acceptance Score≤2 (2)	Acceptance Score=3 (3)	Acceptance Score=4 (4)	Acceptance Score=5 (5)	Average Score (6)	Average Score (7)
Female	0.13	-0.02	-0.02	0.02	0.01	0.09	0.15
	(0.09)	(0.03)	(0.03)	(0.04)	(0.04)	(0.07)	(0.14)
Female × Acceptance score=3							-0.19
							(0.16)
Female × Acceptance score=4							-0.12
							(0.14)
Female × Acceptance score=5							-0.09
							(0.15)
N	853	853	853	853	853	853	853
Year FE	X	X	X	X	X	X	X
Paper controls	X	X	X	X	X	X	X
Referee controls	X	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 referee and paper controls, as specified in equation (2). 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 7: average of technical merit, readability, originality, and relevance scores. In column 7, the female dummy is interacted with the reviewer's acceptance score. The categorical values of reviewer acceptance score are used, with scores ≤ 2 as the omitted category. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix

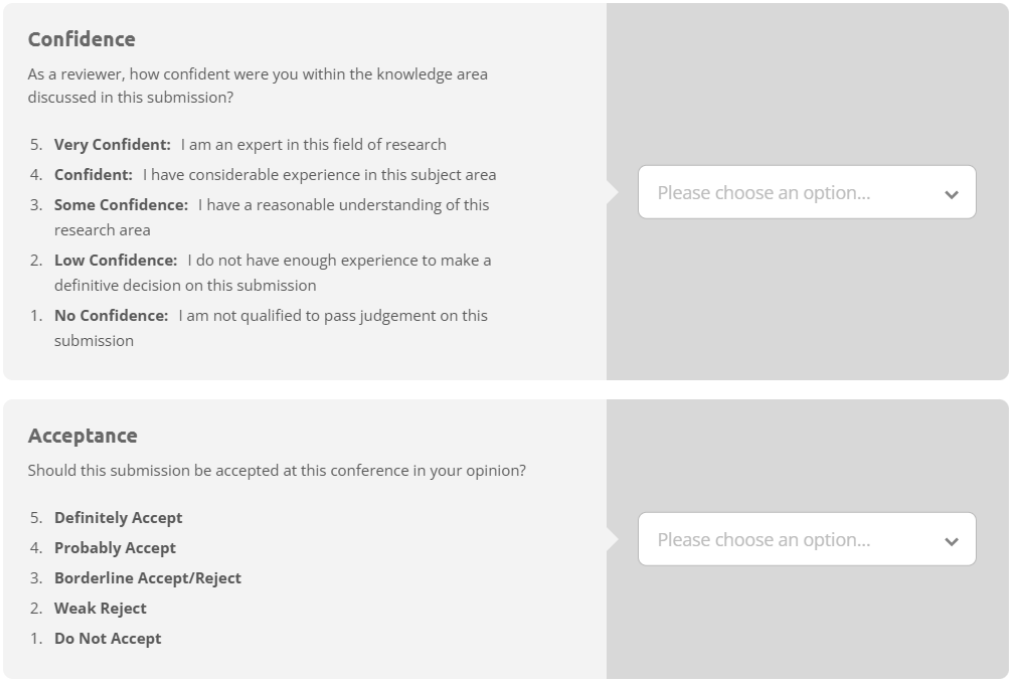
Appendix A Marking Scheme

Figure A.1- Marking Scheme



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

Figure A.2- Confidence and acceptance questions



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

Appendix B LLM Measure of expertise

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

As a robustness check, we use LLM to classify referees into expert or non-expert as described below. First, using ChatGPT, 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

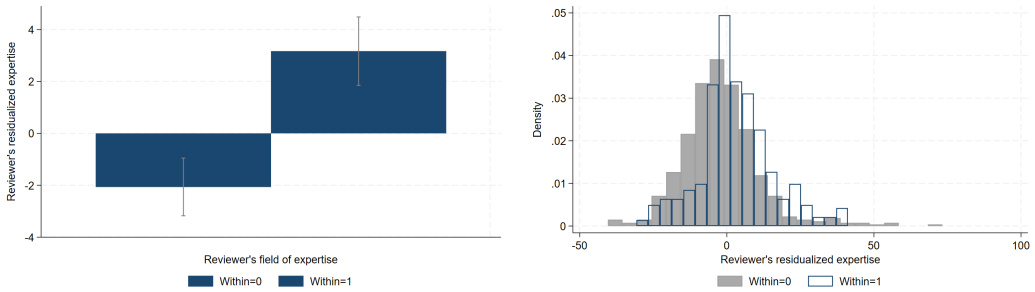
- ...

> 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 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 B.1 displays the distribution of the residuals of objective reviewer's expertise obtained, net of referee 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 referees' expertise scores by subjective expertise, while partially overlapping, are significantly different (Kolmogorov-Smirnov test of difference in distribution p-value=0.00).

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



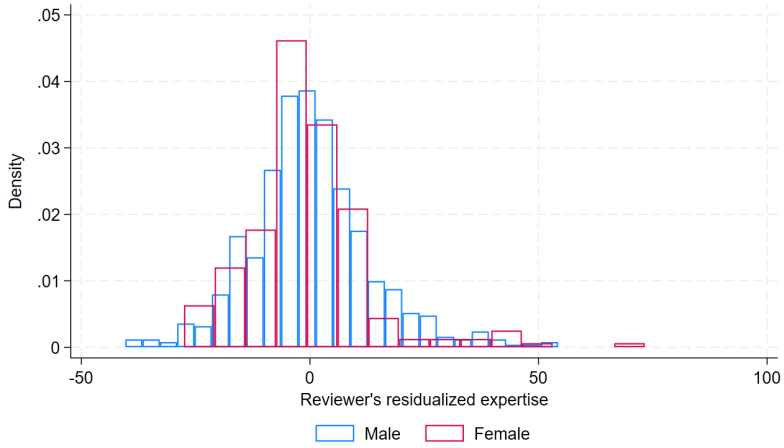
(a) Difference in mean **(b)** Difference in distribution

Notes: Panel (a) displays the average of the residuals of the objective referees' expertise score, net of referee characteristics and year-fixed effects, by referees' subjective field of expertise (measure used in the main analysis). Panel (b) shows the histogram of these residuals by referees' subjective field of expertise. Referee 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 B.2 displays the distribution of residuals of male and female objective reviewers' expertise, net of referee characteristics and year fixed effects. On average, male and female referees 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 3).

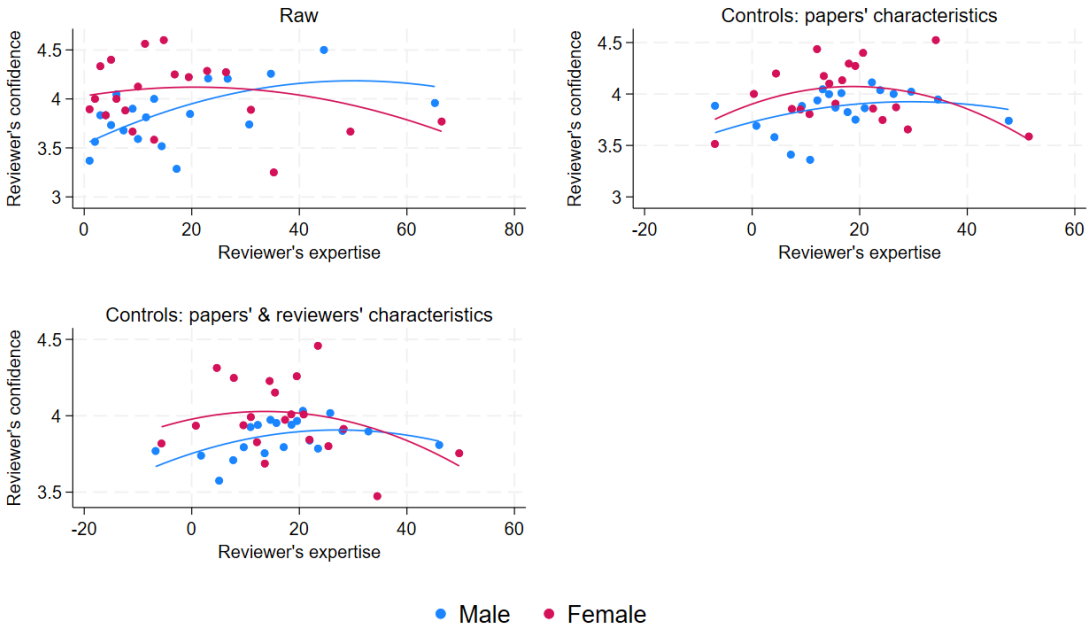
Finally, Figure B.3 plots the correlation of reviewer expertise and confidence. The top left panel displays the raw correlation, the top right panel the correlation controlling for paper characteristics, while the bottom left panel the correlation controlling for paper and reviewer characteristics. Consistent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

Figure B.2- Distribution of expertise



Notes: Histogram of residuals from a regression of the reviewers' expertise on referee 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).

Figure B.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. Referee 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).