



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Confident, but Undervalued: Evidence from the Irish Economic Association Conference

Margaret Samahita and Martina Zanella

TEP Working Paper No. 0325

April 2025

Revision Date: April 2025

Trinity Economics Papers
Department of Economics

CONFIDENT, BUT UNDERVALUED: EVIDENCE FROM THE IRISH ECONOMIC ASSOCIATION CONFERENCE*

Margaret Samahita[†] Martina Zanella[‡]

This paper examines the gender influence gap in an academic setting, focusing on the Irish Economic Association (IEA) Conference review process. Using data from 2017 to 2023, we analyze whether organizers follow the recommendations of male and female reviewers equally and whether any difference can be attributed to a gender gap in the confidence of reviewers. Our findings reveal that organizers' decisions more closely align with male reviewers', particularly when the reviewer's confidence is high and when they have experience in the profession. The influence gap cannot be explained by female reviewers being less confident than males, which is the traditional explanation in the literature. Contrary to expectations, female reviewers report higher confidence than males. We explore potential mechanisms and find suggestive evidence that female reviewers strategically overstate their confidence in anticipation of discriminatory treatment by organizers.

*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.

[†]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 has documented the gender gap in the valuation of expertise: female expertise is often undervalued relative to males'. Many papers have studied this gender gap in the economic profession: in seminars and conferences (Doleac et al., 2021; Hospido and Sanz, 2021; Dupas et al., 2021), recognition of work (Sarsons et al., 2021), citations (Koffi, 2021), teaching evaluations (Mengel et al., 2019; Boring, 2017; Boring and Philippe, 2021), promotions (Lundberg and Stearns, 2019; Baltrunaite et al., 2022; Eberhardt et al., 2023), and the publication process (Card et al., 2020; Hengel and Moon, 2020; Hengel, 2022). One of the explanations traditionally given in the literature is the lack of confidence of women relative to men and female's reluctance to offer opinions (e.g. Jakobsson, 2012; Gordon and Dahl, 2013; Sarsons and Xu, 2021; Sievertsen and Smith, 2022; Coffman et al., 2024; Exley and Kessler, 2022; Exley and Nielsen, 2024). Thus, even when no discriminatory motives are present, women's underconfidence in their own competence, work, and opinion may influence evaluators' judgment.

In this paper, we focus on expertise and evaluations regarding *other people's work*—a crucial and common task in academics' day-to-day work. This is a context where incentives are likely to be different from the settings studied in the other papers, where individuals are asked to evaluate themselves. For women, appearing confident when evaluating others' work may carry lower reputational and identity costs compared to stating confidence in one's own work. Moreover, in this context, stated confidence has consequences for the person whose work is being evaluated, making this a novel and interesting setting to study.

We study organisers' and reviewers' decisions in the Irish Economic Association (IEA) Conference from 2017-2023.¹ We explore the following questions: (i) Do organisers equally value the recommendations of male and female referees? ii) Can any gender influence gap be attributed to a gender gap in referees' confidence? Our empirical strategy exploits a key characteristic of the revision process. For each paper to be reviewed, the referee is asked to provide not only an evaluation of the paper along various dimensions but also a score regarding their confidence in their own expertise. We couple these data with information regarding referees, authors, and papers, collected by looking at CVs and institutional/personal webpages, and Google Scholar.

Our first result is that we find a significant gender influence gap. We compare organizers' acceptance decisions for papers refereed by reviewers of different genders who provide the same evaluation, controlling for an array of characteristics of the paper and the reviewer. We find that when referees are undecided, organisers' decisions more closely align with male referees than female referees. The differential treatment of male and female evaluations also extends to likely accepted papers when reviewers have more experience (seniority) in the profession and when they

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

declare higher confidence.

The influence gap cannot be explained by female referees being less confident than males, which is the traditional explanation in the literature. In an analogous way as for the gender influence gap, we identify the gender gap in confidence by comparing the confidence score provided on similar papers by reviewers of different genders that share otherwise similar individual characteristics. Unlike previous studies, we find that female reviewers declare on average to be *more* confident than male reviewers. This positive gender gap in confidence is stronger the more senior the referee.

One explanation that can reconcile this a-prima-facie surprising result—that females state higher confidence—with the gender influence gap is that female reviewers may be compelled to state higher confidence because they anticipate their expertise more likely to be questioned by the organizers. Four pieces of evidence support this hypothesis of strategic motive. Firstly, we find that the gender confidence gap is driven by papers outside the reviewers’ field of expertise, where female reviewers may perceive a greater likelihood of their recommendations being undervalued. Second, women state higher confidence when they are less familiar with the organizers—a situation where information asymmetry regarding their expertise can trigger organizers to statistically discriminate against them (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 rationale above, both female and male reviewers declare higher confidence when reviewing papers in stereotypically incongruent fields where their expertise is likely to be questioned. Fourth, the gender confidence gap is particularly strong for male-authored papers, a paper characteristic that is also likely to trigger identity considerations and stereotype threat for female reviewers. Finally, we show that this result cannot be explained by selection into seniority or experience in the profession, as we still observe a gender confidence gap for new referees.

Taken together, these 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 consequential? Do organisers’ decisions, by undervaluing female referees, matter for the outcome of the paper under consideration? To provide a tentative answer to this question, we study how reviewers’ evaluations correlate with the paper’s success after the conference. We compare the number of citations accumulated in the years after 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*. This suggests that female referees’ evaluations are more strongly correlated with the actual quality of the paper, pointing to female referees’ expertise being sub-optimally valued.²

²This is in line also with our finding that female reviewers do not appear to be more lenient than males in their

Our paper relates to the rich literature showing that women’s expertise may not be valued as much as their male counterparts. Many examples can be seen in the economics profession (cited above) and other domains (e.g. medicine: de Vaan and Stuart, 2022; Sarsons, 2017; finance: Klein et al., 2021; hiring: Radbruch and Schiprowski, 2023). Our paper documents the gender gap in influence in a novel setting within the economics profession, that of referees’ recommendation to conference organizers. Moreover, it sheds light on a novel consequence of its existence—the dynamic responses of affected agents. Indeed, although there is substantial evidence documenting the existence of differential treatment of women compared to men, fewer studies explore how disadvantaged groups strategically respond to it, particularly outside laboratory 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, though suggestive, provide further insight into one such response strategy.

This paper also enriches our understanding regarding the existence, magnitude, and motives of the gender gap in confidence. Many papers document 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), when asked to express an opinion about themselves.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, 2022; Coffman, 2014). Unlike previous studies measuring confidence when evaluating the individual themselves, their own work, or their own opinion, the men and women in our paper express confidence while evaluating someone else’s work. This is a setting where opportunities and incentives exist to go against societal expectations and use confidence to respond to anticipated discrimination, which is what we document in our paper.

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

evaluations.

³Although 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 A.1). Submission closes in February, after which the conference organizers assign papers to a pool of around fifty reviewers according to their field of research. 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. Reviews 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." A screenshot of the webpage that reviewers are presented with can be found in Appendix Figure B.2.

In 2021, and 2022, the IEA implemented a blind submission system that concealed author names from reviewers. However, reviewers who were also conference organizers could still see all author names. For our analyses, such submissions will be classified as unblinded, along with those from 2017-2019 and 2023.⁵

⁴We 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.

⁵Some 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.

⁶The 2020 conference was canceled due to the Covid-19 pandemic. We also have data from 2016, but no confidence question was asked that year.

⁷2021 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.

⁸47 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

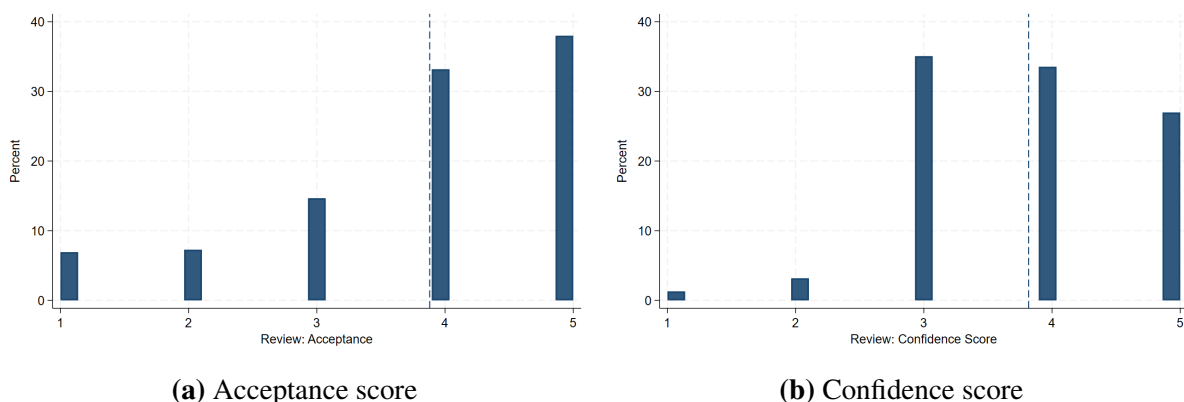
	Male Refs		Female Refs		T-test
	N.	Mean	N.	Mean	Diff. p-value
Within field of expertise	622	0.42	231	0.37	0.15
Within field of expertise (not organizers)	503	0.45	187	0.42	0.47
Male submitter	622	0.71	231	0.69	0.62
Prop. of male authors	622	0.72	231	0.72	0.88
Prop. of authors with PhD	622	0.74	231	0.66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

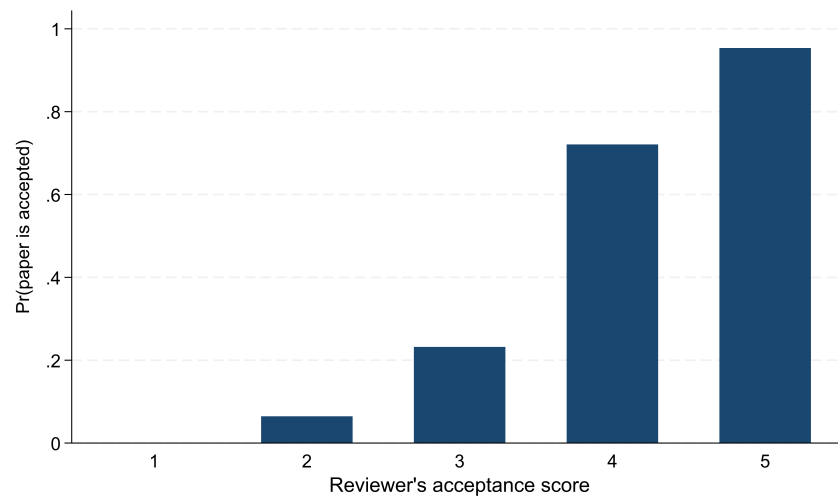


Notes: Histogram of reviewers' acceptance scores (panel a) and confidence scores (panel b). The vertical dashed lines indicate the average scores.

III The Gender Influence Gap

The acceptance score assigned by reviewers significantly influences organizers' decisions, as shown in Figure 2. 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 2- 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).

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.

$$Y_{ipt} = \alpha_t + \sum_{j=2}^{j=5} \beta_j [\mathbb{1}(\text{score} = j)] \times \text{Female}_i + X'_i \gamma + X'_p \delta + \epsilon_{ipt} \quad (1)$$

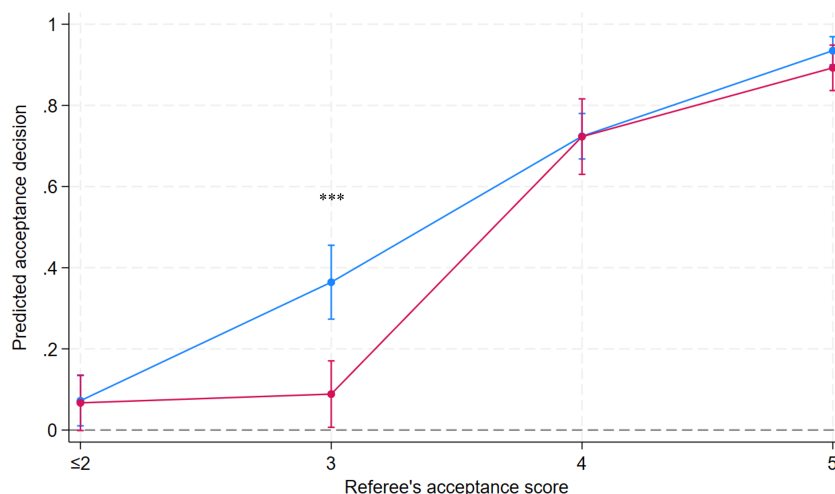
where Y_{ipt} denotes the probability that organizers in year t accept paper p for which reviewer i provided acceptance score j . We control for a set of reviewer-specific controls: affiliation group FEs, N. years post Ph.D. (second order polynomial), Ph.D. institution group FEs, rank FEs, N. of publications (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

Papers that receive similar scores from male and female reviewers do not have the same probability of being accepted. Figure 3 shows the predicted acceptance probability for each referee's

⁹The 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 main results are confirmed if we replicate the analysis by excluding these observations, as shown in Tables A.2 and 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.

acceptance score, net of paper and referee characteristics. While there is no gender gap in papers' success—as measured by its 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 A.2 in Appendix.

Figure 3- Gender gap in the probability of the paper being accepted

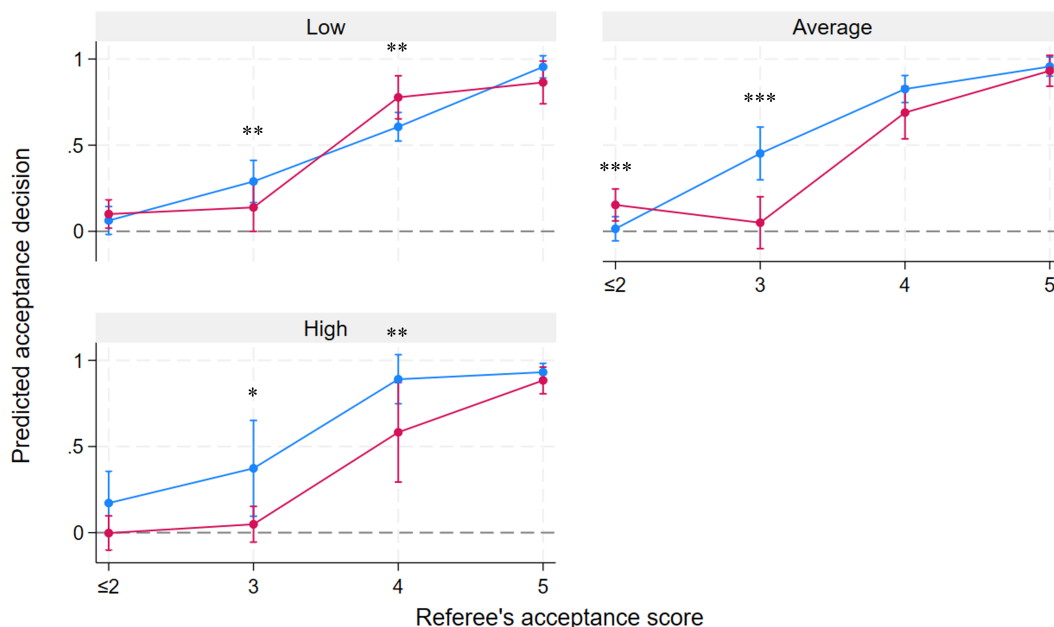


Notes: Average predicted probability that the paper is accepted by the organizers, plotted against reviewer's acceptance decisions, net of referee and paper characteristics as in equation (1). The figure is generated using Stata's *marginsplot* command. Stars indicate the significance of the gender gap 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 4 (estimates provided in the Appendix Table A.3). 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 A.6 in the Appendix 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 4- Gender gap in the probability of the paper being accepted - by confidence



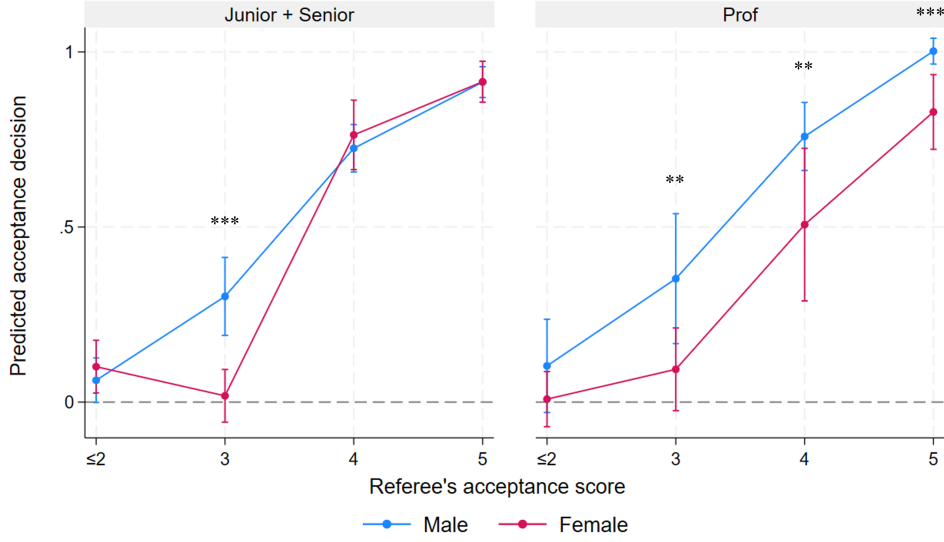
Notes: Average predicted probability that the paper is accepted by the organizers, plotted against reviewer's acceptance decisions, net of referee and paper characteristics as in equation (1). Each graph displays results for a subsample of referees with different confidence: reviewers whose confidence 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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Interestingly, the gender influence gap is particularly strong for Professors, contrary to the literature which suggests that the gender gap in 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, 2024).¹⁰ As shown in Figure 5, despite their experience in the profession, female professors' recommendations are less likely to be followed than male professors across all acceptance scores.¹¹

¹⁰For 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 (2024), 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.

¹¹The gender gap becomes less precisely estimated when we restrict the sample to Professors as the number of observations falls to 254.

Figure 5- 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 (1). 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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

IV The Gender Gap in Confidence

Can the observed gender influence gap be explained by women declaring lower confidence in their evaluations compared to men? This is a common explanation in the literature for why expert opinions are valued differently based on the expert's gender. Indeed, even when no discriminatory motives are present, women's underconfidence may influence evaluators' judgment. For example, Exley and Nielsen (2024) shows that, while evaluators may correctly anticipate the confidence gap, they do not sufficiently update their posterior beliefs and thus form overly pessimistic beliefs about women's abilities. This lack of updating is attributed to a lack of attention paid to the confidence gap.

To identify the 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:

$$Y_{ipt} = \alpha_t + \beta \times Female_i + X'_i \gamma + X'_p \delta + \epsilon_{ipt} \quad (2)$$

where Y_{ipt} denotes the score reviewer i gave to paper p in year t . We control for the same set of

reviewer-specific and paper-specific characteristics as described for equation (1).

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 A.1. 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.

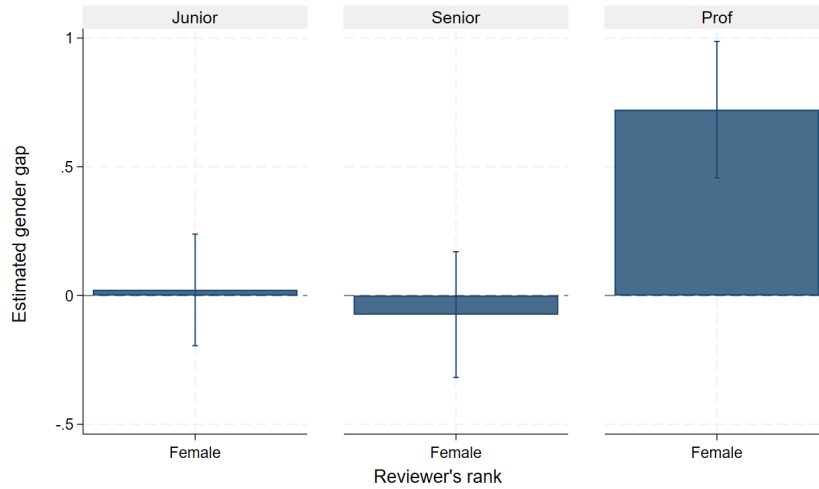
Table 4- Gender gap in declared confidence

	Reviewers' confidence score {1;5}				
	(1)	(2)	(3)	(4)	(5)
Female	0.19*** (0.07)	0.22*** (0.08)	0.19** (0.08)	0.18** (0.08)	0.21** (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 (2). 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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Given the earlier result that the gender gap in the valuation of expertise is stronger among professors, one natural question to ask is whether female professors are less confident than their male counterparts. We therefore check for heterogeneity in the gender confidence gap by seniority in the profession. The results obtained from estimating equation (2), including interactions of the referee's gender with their rank, are shown in Figure 6. The average estimated difference in confidence is driven by reviewers who are Professors or equivalent, while the positive gender effect is not present for junior economists. Table A.7 in the Appendix provides the coefficients behind Figure 6.

Figure 6- 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 (2). 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.

Taken together, the results of the influence and confidence gaps indicate that despite declaring significantly higher confidence, female reviewers' evaluations are followed less closely by the organizers. This is particularly evident for female Professors, who display the biggest (positive) gender gap in confidence, and the biggest (negative) gender influence gap.

V Potential Explanations

What can account for this surprising positive gender gap in confidence? Can this somehow be reconciled with the finding that organizers undervalue female expertise? 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 in two steps. First, we provide 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. Second, we argue that the positive gender gap cannot be reconciled with alternative explanations, such as selection into professorship or accumulated experience.

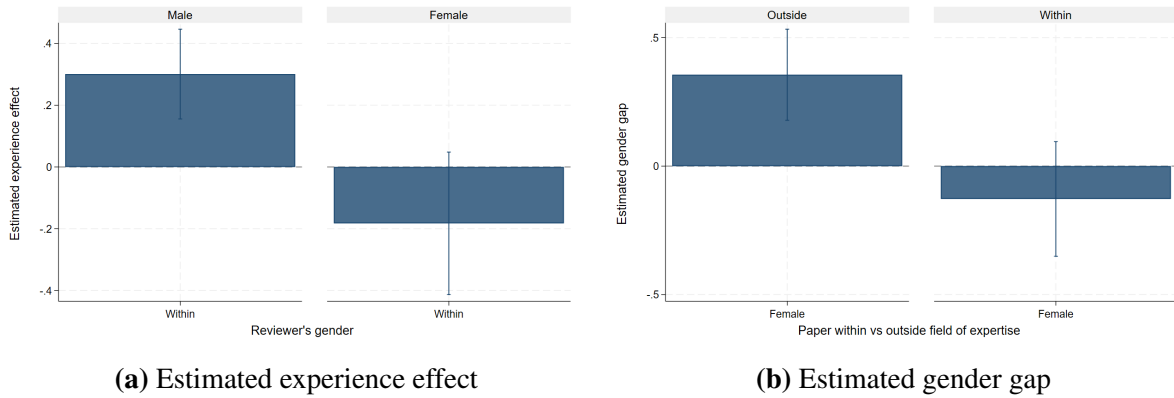
VA Strategic motive

Field of expertise 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 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.¹² We then estimate specification (2) 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 7 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 A.8 in the Appendix provides the complete set of estimates Figure 7 refers to.

Figure 7- The role of expertise



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 (2). 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.

¹²Table 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 C we provide a different measure of field expertise using LLM, our results are robust to this alternative definition.

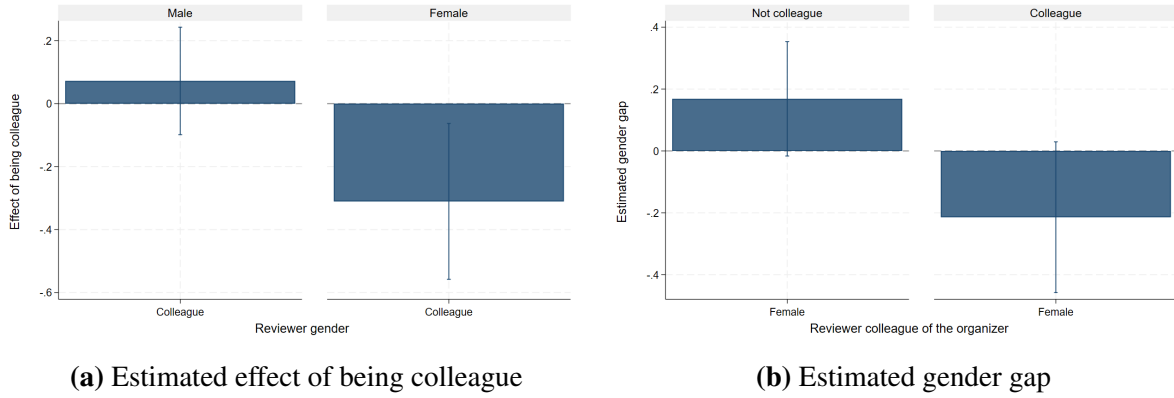
This result is again surprising. Other papers studying the Economics profession find that senior female economists are less confident than men when responding to questions or providing opinions about the economy and government policy outside their field of expertise (Sarsons and Xu, 2021; Sievertsen and Smith, 2022). 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 benefit of greater influence that can derive from declaring higher confidence.

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 8 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 (2). 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 8, panel b). Table A.9 in the Appendix 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 8- The role of familiarity to organizers



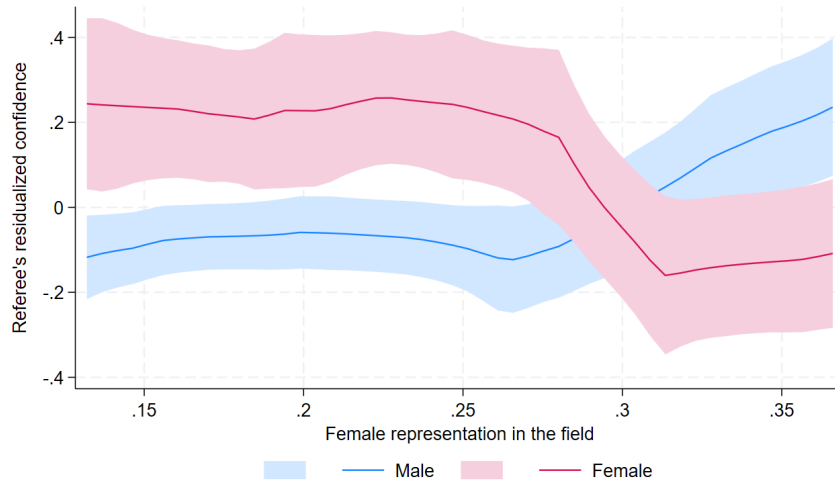
Notes: Bar graph that displays the estimated effect of being a colleague of the organizers for men and women (panel a) and the estimated gender confidence gap (female minus male) for non-colleagues and colleagues of organisers (panel b), net of paper and referee controls as specified in equation (2). 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 the possibility of their expertise being undervalued, 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 A.10 in the Online Appendix.

Figure 9 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 (2). 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 A.11.

¹³The 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 9- The role of gender stereotype of paper topic



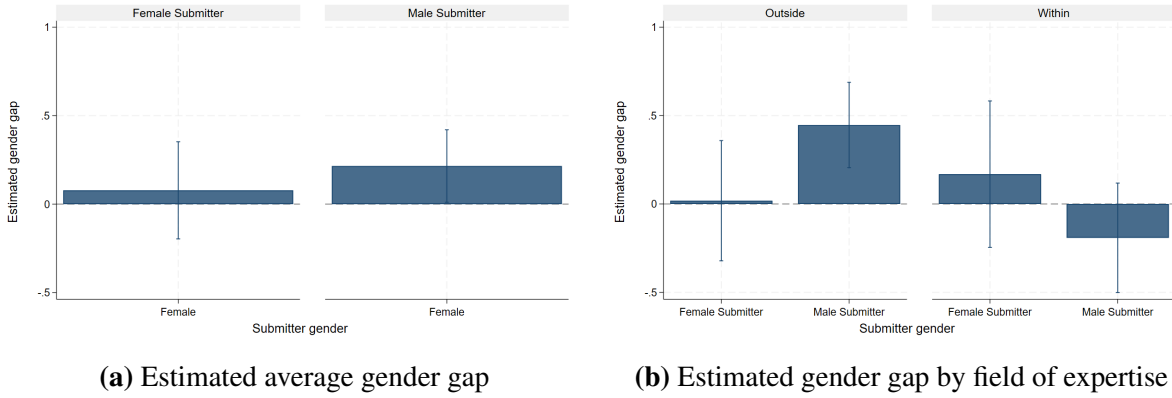
Notes: The graph displays the correlation between the share of women working in the paper's primary field (x axis) and the confidence score provided by male (blue line) and female (red line) reviewers (y axis) for each paper. The red and blue lines display the smoothed values of a kernel local polynomial regression of the residuals from a regression of the reviewers' confidence score on paper and referee characteristics, as described in equation (2). Female representation in each field is constructed using the statistics regarding the share of women working in each field in Economics provided by IDEAS. The categorization can be found in Table A.10 in Appendix.

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 10 shows the gender confidence gap by the gender of the submitter, for unblinded submissions, net of paper and reviewers' characteristics as described in equation (2). 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 A.12 in the Appendix 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.

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

Figure 10- The role of the submitter's gender



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 (2). 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—a rational expectation which is in turn consistent with the evidence we presented in Section III.

VB Alternative mechanisms

Selection into seniority in the profession Given the gender bias observed in the economics profession (Doleac et al., 2021; Dupas et al., 2021; Eberhardt et al., 2023; Baltrunaite et al., 2022; Sarsons et al., 2021; Mengel et al., 2019; Boring, 2017; Boring and Philippe, 2021), those who make it as a female Professor may have been self-selected by their high confidence.

Table A.13 in the Appendix provides evidence that selection into senior positions in the profession alone cannot explain why we observe the gender confidence gap. Indeed, also for Professors, the positive gender gap is driven by papers outside the field of expertise (panel b). Female professors do not display significantly higher confidence compared to male professors when the paper is within the field of expertise. This is due to the fact that female Professors, like all other women, do not declare higher confidence when reviewing papers within the field of expertise (panel a). If anything, they declare lower confidence (not significant).

Experience A second alternative explanation for the observed positive gender gap in confidence is that females who make it to the top have accumulated experience as reviewer, thus increasing

their confidence.

Figure A.2 in Appendix shows that rank in the profession¹⁵ is indeed positively correlated with experience in reviewing papers, as proxied by the number of papers reviewed in previous conferences, and the number of papers previously reviewed is strongly positively correlated with the confidence that reviewers declare themselves. However, while experience in reviewing may contribute to the development of strategic behavior—as women learn to anticipate that their expertise might be discounted, for example—it cannot fully explain the observed gap. In Table A.14 we estimate the gender confidence gap for different levels of experience in reviewing. In particular, we distinguish between reviewers with zero experience (rookies) and reviewers who reviewed at least one paper before. We find that the gender gap in confidence is driven by reviewers with previous experience. However, rookie referees also display a positive gender gap when reviewing papers outside the field of expertise. This confirms that something beyond experience is behind the estimated positive gender gap in confidence.

VI Is this consequential? Paper outcomes

Do organisers’ tendencies to undervalue female referees matter for the outcome of the paper under consideration? 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 11. 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.

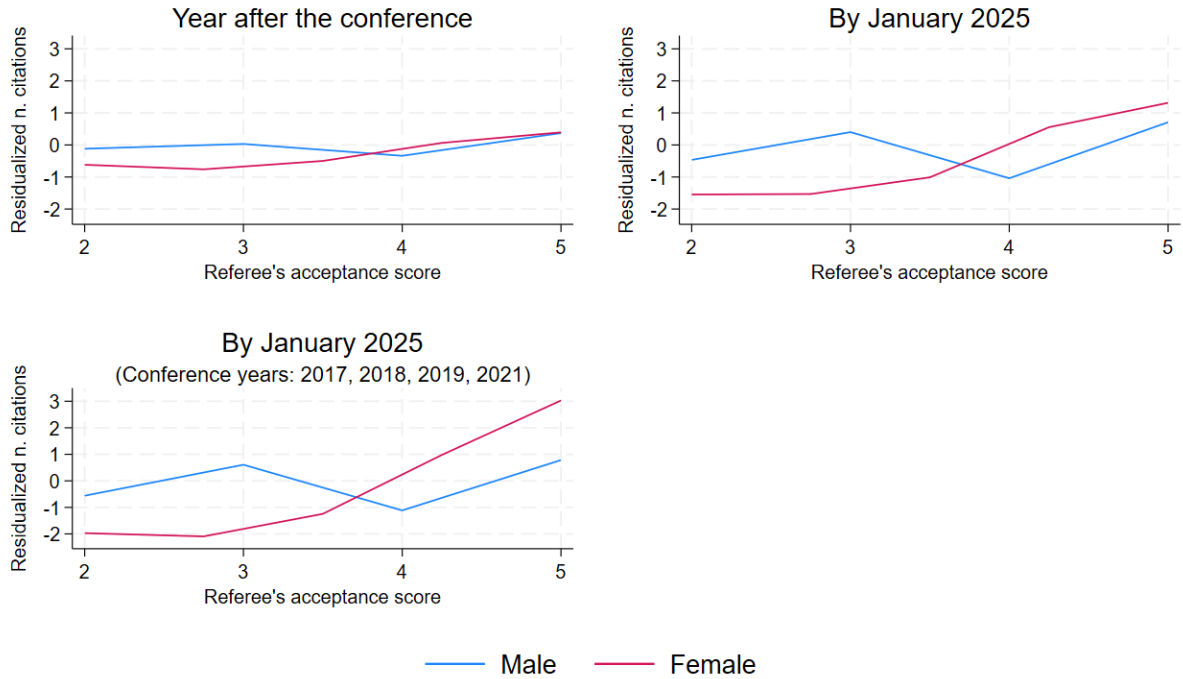
Female reviewers’ evaluations appear to correlate more positively with paper success. Notably, the graphs show an overall steeper slope for female reviewers compared to males. For low-acceptance-score papers, female-refereed papers accumulate a lower number of citations. The gender gap reverses for high-acceptance-score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

¹⁵1 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.

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 11- 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 (1)) 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.

VII Conclusion

In this paper, we document the existence of a gender gap in influence and confidence in the context of evaluations regarding other people's work using data from the Irish Economic Association (IEA) Conference from 2017-2023. We find that female referees have lower influence than their male counterparts as organisers' acceptance rates are much more closely aligned with male referees' recommendations. This gender influence gap cannot be accounted for by the existence of a gender confidence gap, where female referees declare to be less confident than males. In fact, we find a

gender confidence gap in the *opposite* direction: female reviewers declare to be more confident than their male counterparts. The evidence suggests that the observed patterns can be reconciled with female reviewers strategically overstating their confidence in anticipation of discriminatory treatment by the organizers.

We can speculate that this represents an inefficient equilibrium. Firstly, referees' strategic response, while potentially optimal in a static setting, runs the risk of triggering a vicious cycle that perpetuates and amplifies the initial differential treatment by organizers. Organizers may increasingly discount female reviewers' recommendations not only due to any potential gender bias, but also because they correctly anticipate female referees being strategic in overstating their confidence. 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.

The paper studies a specific population – Irish economists – and context – conference acceptance decisions. Nonetheless, the patterns documented here likely extend to other settings where women, or other underrepresented groups, are tasked with evaluating others' work and may anticipate a higher likelihood of discriminatory treatment. In such environments, strategic behavior such as overstating confidence can emerge as a rational response to minimize the consequences of anticipated discrimination. The act of assessing others is widespread across many domains. Aside from academic reviewing, other examples include hiring and promotion decisions in the workplace, assessing candidates for loan applications, or providing a reference for a potential tenant. 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 Europe¹⁶ and in other influential spheres such as politics and corporate governance.¹⁷ In addition, the relatively small size of the Irish economics community likely fosters close professional networks and high interpersonal familiarity, which may attenuate the dynamics identified in this study.¹⁸ 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.

¹⁶This 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.

¹⁷The 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](#).

¹⁸The *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.

References

- 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.
- Card, D., S. DellaVigna, P. Funk, and N. Iriberry (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. 10th miami behavioral finance conference, available at ssrn: <https://ssrn.com/abstract=3318245> or <http://dx.doi.org/10.2139/ssrn.3318245>.
- 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. Zwillling (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. Punčohář (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. and S. Smith (2022). Male and female voices in economics. Centre for Economic Policy Research Discussion Paper No. 17166.
- Sievertsen, H. and S. Smith (2024). Do female experts face an authority gap? Evidence from economics. IZA Discussion Paper No. 17029.
- Zanella, M. (2024, January). Stereotypical Selection. Trinity Economics Papers tep0224, Trinity College Dublin, Department of Economics.

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

Appendix A Additional Tables and Figures

Table A.1- Origin of Submitters

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

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

Table A.2- The gender influence gap

	Organizers' acceptance decision {0; 1}			
	(1)	(2)	(3)	(4)
Referee's acceptance score	0.26*** (0.01)	0.25*** (0.01)		
Female	-0.07** (0.03)	-0.19*** (0.07)	-0.01 (0.04)	0.02 (0.04)
Female $\times$ Referee's acceptance score		0.03* (0.02)		
Referee's acceptance score=3			0.25*** (0.06)	0.31*** (0.07)
Referee's acceptance score=4			0.67*** (0.04)	0.68*** (0.05)
Referee's acceptance score=5			0.88*** (0.04)	0.89*** (0.04)
Referee's acceptance score=3 $\times$ Female			-0.26*** (0.07)	-0.33*** (0.08)
Referee's acceptance score=4 $\times$ Female			0.00 (0.07)	-0.00 (0.07)
Referee's acceptance score=5 $\times$ 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 [0.00]	-0.32 [0.00]
Gender gap if Referee's acceptance score=4			-0.01 [0.92]	0.16 [0.79]
Gender gap if Referee's acceptance score=5			-0.05 [0.11]	-0.03 [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 (1). 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 A.3- 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 A.5 in Panels A-C and Table A.4 in Panels D and E. P-values in square brackets.

Table A.4- 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 (1), interacting the female x score interaction with a Professor dummy. Professor is a dummy equal to one for rank 5. The categorical values of reviewer acceptance score are used, with scores ≤ 2 as the omitted category. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5- The gender influence gap by declared confidence

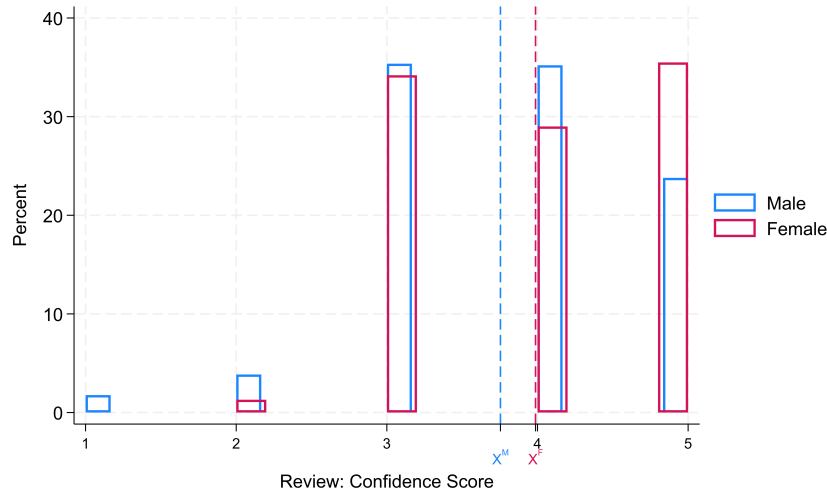
	Organizers' acceptance decision {0, 1} (1)
Referee's acceptance score=3	0.19** (0.08)
Referee's acceptance score=4	0.54*** (0.06)
Referee's acceptance score=5	0.89*** (0.06)
Female	0.01 (0.05)
Referee's acceptance score=3 × Female	-0.21** (0.09)
Referee's acceptance score=4 × Female	0.16* (0.09)
Referee's 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 (1), interacting the female x score interaction with a dummy equal to one if the reviewer declares average confidence, and a dummy equal to one if the reviewer declares high confidence. Low, average, and high confidence correspond to confidence scores less than or equal to 3, equal to 4, and equal to 5, respectively. The categorical values of reviewer acceptance score are used, with scores ≤ 2 as the omitted category. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6- 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 (0.09)	-0.02 (0.03)	-0.02 (0.03)	0.02 (0.04)	0.01 (0.04)	0.09 (0.07)	0.15 (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 (1). The outcome variable is the reviewers' acceptance score ($\{1,5\}$) in column 1; a dummy equal to one for reviewers' acceptance score ≤ 2 , equal to 3, 4, and 5 in columns 2-5, respectively; and the reviewers' average score in column 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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.1- Reviewers' confidence scores by gender

Notes: Histogram of the reviewers' confidence scores by gender. The vertical lines indicate the average confidence scores for the two groups.

Table A.7- Gender gap in confidence: the role of rank

	(1)	(2)	(3)	(4)
Female	0.07 (0.10)	0.10 (0.11)	0.02 (0.11)	0.05 (0.12)
Female x Senior	-0.03 (0.15)	-0.11 (0.16)	-0.10 (0.16)	-0.04 (0.18)
Female x Prof	0.51*** (0.16)	0.59*** (0.17)	0.80*** (0.17)	0.68*** (0.18)
Senior	0.12 (0.10)	0.14 (0.09)	-0.05 (0.10)	-0.09 (0.11)
Prof	0.36*** (0.10)	0.35*** (0.10)	0.17 (0.11)	0.21* (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 [0.71]	-0.01 [0.96]	-0.07 [0.55]	0.01 [0.96]
Gender gap if Prof	0.58 [0.00]	0.69 [0.00]	0.82 [0.00]	0.73 [0.00]

Note: This table reports the estimated gender gap in confidence by reviewer's rank. The regression specification is given by equation (2), 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 A.8- Gender gap in confidence: the role of expertise

	(1)	(2)	(3)	(4)
Female	0.38*** (0.09)	0.42*** (0.10)	0.36*** (0.09)	0.41*** (0.10)
Female x Within	-0.48*** (0.14)	-0.55*** (0.15)	-0.48*** (0.14)	-0.52*** (0.15)
Within	0.33*** (0.07)	0.31*** (0.08)	0.30*** (0.07)	0.28*** (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 [0.34]	-0.13 [0.26]	-0.13 [0.26]	-0.11 [0.38]
Expertise effect if Female=1	-0.15 [0.19]	-0.24 [0.06]	-0.18 [0.12]	-0.24 [0.07]

Note: This table reports the estimated gender gap in confidence by reviewer's expertise. The regression specification is given by equation (2), 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 A.9- 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 $\times$ Colleague	-0.38**	-0.44**
	(0.15)	(0.20)
Within	0.14**	0.24***
	(0.07)	(0.09)
Female $\times$ Within		-0.53***
		(0.18)
Within $\times$ Colleague		0.15
		(0.17)
Female $\times$ Within $\times$ 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 (2), 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 (2). 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 A.10- 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 A.11- Gender gap in confidence: the role of the paper's field stereotypicality

	(1)
Female	1.07***
	(0.21)
Female $\times$ 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 (2), interacting the female dummy with the female representation in each field - Female Share ($[0;1]$), constructed using the statistics regarding the share of women working in each field in Economics provided by IDEAS. The categorization can be found in Table A.10 in Appendix. Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.12- Gender gap in confidence: authors' gender

	Male submitter		Prop. male authors	
	(1)	(2)	(3)	(4)
Female	0.08 (0.14)	0.02 (0.17)	0.16 (0.17)	0.02 (0.22)
Male Authors	-0.14 (0.09)	-0.30*** (0.11)	-0.06 (0.11)	-0.37** (0.14)
Male Authors $\times$ Female	0.14 (0.16)	0.43** (0.20)	0.02 (0.19)	0.41 (0.26)
Within		0.02 (0.14)		-0.20 (0.18)
Within $\times$ Female		0.15 (0.26)		0.28 (0.32)
Within $\times$ Male Authors		0.39*** (0.16)		0.68*** (0.21)
Within $\times$ Male Authors $\times$ Female		-0.79*** (0.31)		-0.96*** (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 [0.04]		0.18 [0.09]	
Gender gap if Within=0, Male Authors		0.45 [0.00]		0.43 [0.00]
Gender gap if Within=1, Female Authors		0.17 [0.42]		0.31 [0.21]
Gender gap if Within=1, Male Authors		-0.19 [0.23]		-0.25 [0.12]

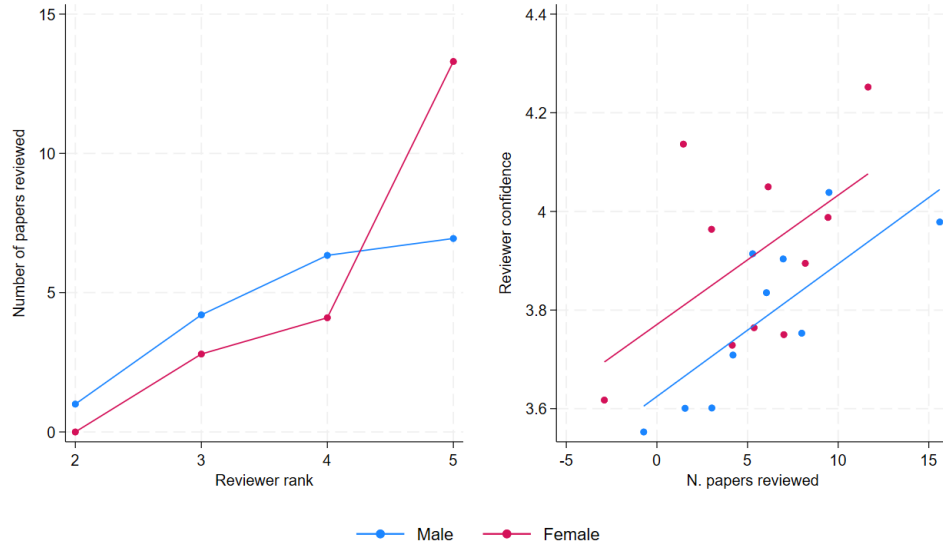
Note: This table reports the estimated gender gap in confidence by authors' gender. The regression specification in columns 1 and 3 is given by equation (2), interacting the female dummy with a dummy equal to one if the submitter of the paper is male (columns 1), and the proportion of authors who are 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 (2). 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 A.13- Gender gap in confidence: the role of expertise by rank

	(1)
Female	0.16 (0.11)
Prof	0.27** (0.12)
Female $\times$ Prof	0.66*** (0.17)
Within	0.39*** (0.09)
Female $\times$ Within	-0.45*** (0.16)
Prof $\times$ Within	-0.25* (0.14)
Female $\times$ Prof $\times$ 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 (2). Robust standard errors in parentheses; P-values in square brackets; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.2- 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 A.14- Gender gap in declared confidence: the role of experience

	(1)	(2)
Female	0.26*** (0.09)	0.47*** (0.11)
Female × Rookie	-0.19 (0.15)	-0.21 (0.18)
Female × Within		-0.55*** (0.16)
Female × Within × Rookie		0.05 (0.28)
Within	0.16** (0.06)	0.30*** (0.09)
Rookie	-0.14 (0.09)	-0.17 (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 (2). Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B Marking Scheme

Figure B.1- Marking Scheme

Marking Scheme

Technical Merit

How solid is the presented work? Is the evaluation methodology appropriate? Does the data seem accurate? Are there any fatal flaws in underlying assumptions?

1. **Unacceptable:** Submission has serious errors in approach that invalidate the results, or clearly erroneous data.
2. **Poor:** Methodology is unclear, data may have major errors (but unclear), questionable assumptions.
3. **Acceptable:** Only minor flaws in method/data.
4. **Good:** Seems technically sound.
5. **Excellent:** Exceptionally thorough/accurate in methodology and results.

12345

Readability

How easy is it to understand the submission? Factors that can affect readability include writing style, grammar, spelling, over-use (or under-

12345

Relevance

How appropriate is this submission for this conference? Sometimes even good submissions are better suited to other locations. If you mark this

12345

Originality

Will attendees learn something that they didn't already know from this submission?

12345

Format

While uploading a submission, authors indicated a preference for either an oral or poster presentation (shown above). If accepted, which

Please choose an option...

Confidence

As a reviewer, how confident were you within the knowledge area discussed in this submission?

Please choose an option...

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

Figure B.2- Confidence and acceptance questions

The image shows two sections of a form. The top section is titled 'Confidence' and asks 'As a reviewer, how confident were you within the knowledge area discussed in this submission?'. It lists five options: 5. Very Confident: I am an expert in this field of research; 4. Confident: I have considerable experience in this subject area; 3. Some Confidence: I have a reasonable understanding of this research area; 2. Low Confidence: I do not have enough experience to make a definitive decision on this submission; 1. No Confidence: I am not qualified to pass judgement on this submission. To the right of the list is a dropdown menu with the text 'Please choose an option...' and a downward arrow. The bottom section is titled 'Acceptance' and asks 'Should this submission be accepted at this conference in your opinion?'. It lists five options: 5. Definitely Accept; 4. Probably Accept; 3. Borderline Accept/Reject; 2. Weak Reject; 1. Do Not Accept. To the right of the list is another dropdown menu with the text 'Please choose an option...' and a downward arrow.

Confidence
As a reviewer, how confident were you within the knowledge area discussed in this submission?

5. **Very Confident:** I am an expert in this field of research
4. **Confident:** I have considerable experience in this subject area
3. **Some Confidence:** I have a reasonable understanding of this research area
2. **Low Confidence:** I do not have enough experience to make a definitive decision on this submission
1. **No Confidence:** I am not qualified to pass judgement on this submission

Please choose an option...

Acceptance
Should this submission be accepted at this conference in your opinion?

5. **Definitely Accept**
4. **Probably Accept**
3. **Borderline Accept/Reject**
2. **Weak Reject**
1. **Do Not Accept**

Please choose an option...

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

Appendix C 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

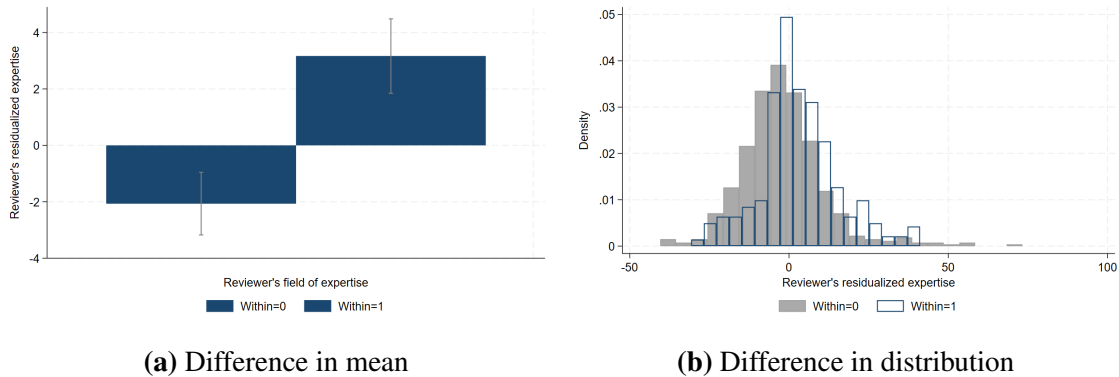
• ...

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 C.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).

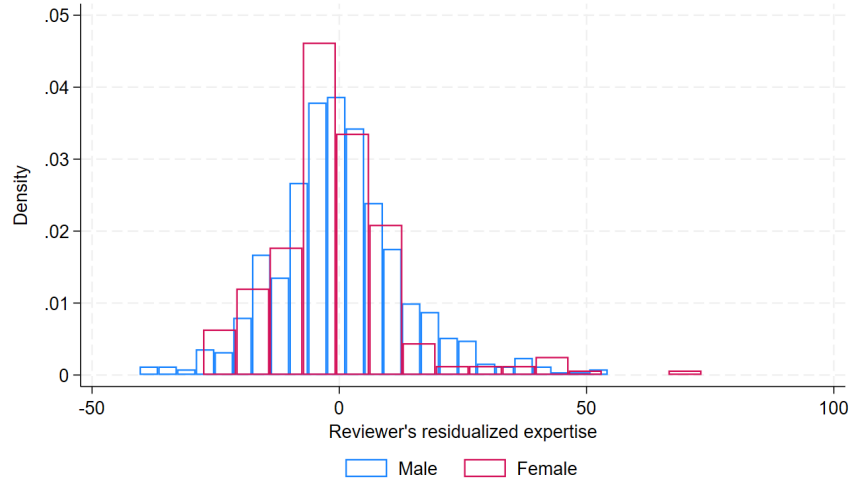
Figure C.1- Distribution of expertise by field of expertise



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 C.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).

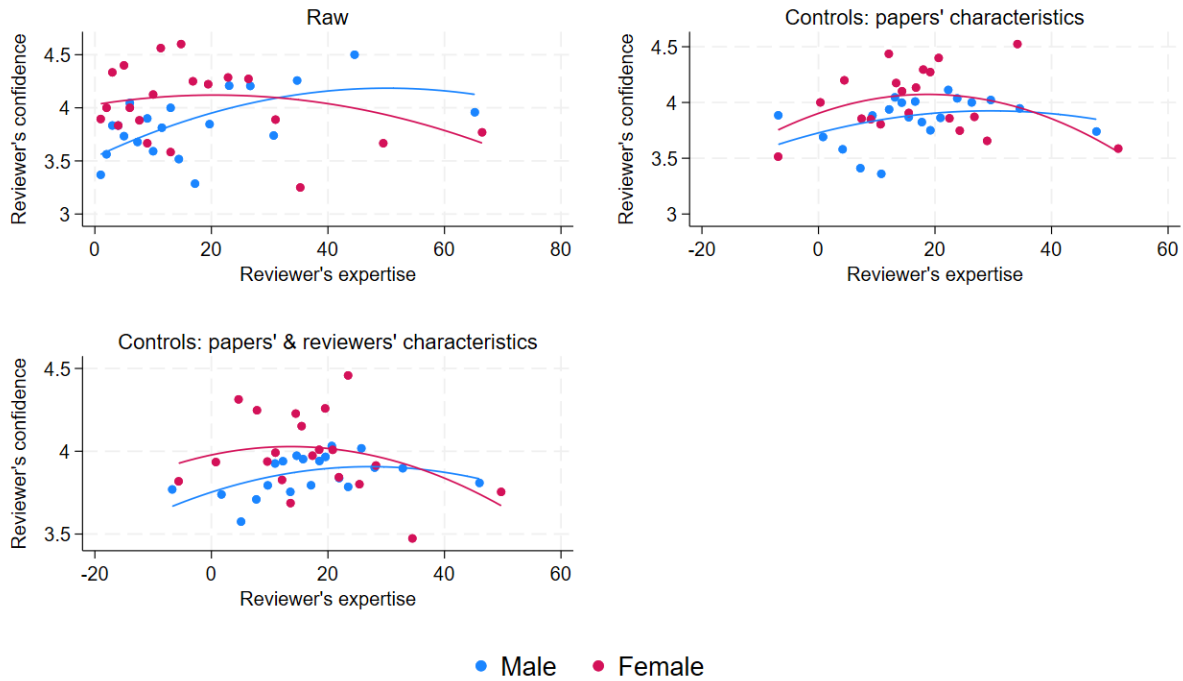
Figure C.2- Distribution of expertise



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).

Finally, Figure C.3 plots the correlation of reviewer expertise and confidence. The top left panel displays the raw correlation, the top right panel the correlation controlling for paper characteristics, while the bottom left panel the correlation controlling for paper and reviewer characteristics. Consistent with earlier findings, we observe that male 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 C.3- Correlation of reviewer expertise and confidence



Notes: Binned scatter plots displaying the correlation between reviewers' confidence scores and reviewers' objective expertise. The top left panel displays the raw correlation, the top right panel the correlation controlling for paper characteristics, while the bottom left panel the correlation controlling for paper and reviewer characteristics. 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 (1).