Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Legibility Gap: How Gender Equity Interventions Redistribute Recognition Across Cultures

Dashun Wang

dashun.wang@kellogg.northwestern.edu

Northwestern University https://orcid.org/0000-0002-7054-2206

Binglu Wang

Kellogg School of Management, Northwestern University https://orcid.org/0000-0002-0013-1276

Jose Cervantez

University of Pennsylvania

Jiahui Xue

Kellogg School of Management, Northwestern University

Katherine Milkman

The Wharton School, University of Pennsylvania https://orcid.org/0000-0002-9706-4830

Article

Keywords:

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11363709/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

The Legibility Gap: How Gender Equity Interventions Redistribute Recognition Across Cultures

Binglu Wang[1,2,3,4], Jose Cervantez[5], Jiahui Xue[1,2], Katherine Milkman[5], and Dashun Wang[1,2,3,4,6*]

[1] Kellogg School of Management, Northwestern University; Evanston, IL, USA
[2] Center for Science of Science and Innovation, Northwestern University; Evanston, IL, USA
[3] Northwestern Innovation Institute, Northwestern University; Evanston, IL, USA
[4] Ryan Institute on Complexity, Northwestern University; Evanston, IL, USA
[5] The Wharton School of the University of Pennsylvania; Philadelphia, PA, USA
[6] McCormick School of Engineering, Northwestern University, Evanston, IL, USA

* Corresponding author. Email: dashun.wang@kellogg.northwestern.edu

Abstract:

Efforts to promote gender equity in science increasingly rely on name-based inference to quantify representation and guide policy and behavior. Yet linguistic cues that signal gender vary across cultures and are often obscured when names are transliterated into English. Here we identify a pattern we call the "legibility gap": when gender is inferred from names, equity interventions systematically benefit women whose names signal gender while bypassing those whose names lose such cues in translation. Using both observational and experimental evidence, we show how this gap reshapes recognition in science. Analyzing citation diversity statements—an emerging practice in which authors report the algorithmically estimated gender composition of their reference lists—we find that papers that include this practice cite women more frequently, but the gains accrue almost entirely to authors with gender-signaling Western names. By contrast, women whose names lose gender cues in English transliteration, predominantly those with East Asian names, receive fewer citations in these same papers. Two preregistered experiments ($N = 2{,}250$) corroborate this pattern and identify its mechanism: linguistic legibility, not cultural unfamiliarity, determines who is recognized as a woman and who benefits from policies designed to support women in science. Overall, these findings expose a previously unrecognized layer of inequity embedded in global equity infrastructures. As science becomes increasingly global and equity efforts increasingly algorithmic, the legibility gap reveals how uneven identity recognition reshapes fairness. In global systems of recognition, equity depends not only on whether policies are effective on average, but also on whether they are equitable across cultures.

1. Introduction

Efforts to promote women in science have accelerated in recent years, in part due to growing awareness of persistent disparities in hiring, credit, promotion, and visibility [1-17]. But promoting equity first requires a way to identify women in the data. Names have long shaped how women are recognized and evaluated [18,19]: women authors have adopted male pseudonyms to evade bias [20], and controlled experiments show that identical résumés are judged differently depending on whether a male or female name is attached [3,21-23]. Today, these dynamics operate at a far greater scale. Because verified demographic data are often unavailable, incomplete, or protected for privacy reasons [24], researchers and institutions increasingly rely on name-based inference—by algorithms and by human readers—to identify gender and monitor inclusion efforts [5,6,8,9,12-16].

Yet names are not equally legible across languages. In many naming systems, gender cues are embedded in spelling, characters, or pronunciation that disappear when names are transliterated into English [25-27], the dominant language of science [28,29]. As a result, some women's names remain readily recognizable as female while others do not, even when gender is clear in the original language. When equity efforts depend on inferred gender, this uneven legibility may create a consequential asymmetry: the very systems designed to support women may end up supporting only those whose names are easiest to classify as women.

While name-based inference rarely determines an outcome on its own, its significance lies upstream at the opportunity layer where equity efforts typically intervene, by shaping who is visible in a candidate or expert pool, how representation is measured, which disparities are diagnosed, and where corrective attention is directed (Table S1). This infrastructure now operates across hiring and workforce audits, official statistics, expert directories, and large-scale analyses of science [5,6,8,9,12-16] (SI Section S1). Recent work, including our validation across 4.5 million scientists (see SI Section S1.1), shows that these systems perform substantially worse on transliterated East Asian names than on Western ones [25,26,30]. Uneven legibility may therefore be more than a source of measurement error. By shaping who is counted, surfaced, and targeted by equity efforts, it may redistribute opportunity toward those whose identities are easiest for the system to see.

In this paper, we test that possibility using both observational and experimental evidence. We first examine citation diversity statements, an emerging practice in which authors report the gender composition of their reference lists, creating a natural setting in which inferred gender shapes behavior [31,32]. We then present two preregistered experiments involving participants from five cultures that test whether interventions designed to support women in science disproportionately benefit those whose names remain legible as female in English. Across both settings, we find that interventions intended to promote gender equity help women overall but disproportionately benefit women with Western names that clearly signal gender, while bypassing women whose names lose gender cues in transliteration. We term this pattern the "legibility gap."

2. Study 1: Citation diversity statements and the redistribution of recognition

To test whether equity efforts built on inferred gender may redistribute recognition unevenly across cultures, we begin by analyzing citation diversity statements (CDSs) in science. CDSs make inferred gender behaviorally consequential: authors are encouraged to calculate the gender composition of the scholars they cite using an automated classification tool and then disclose that composition in their published paper [31,32]. If gender cues are not equally legible across names, then a practice designed to increase citations to women may not benefit all women equally. Instead, it may disproportionately reward those whose names are easiest for the system to recognize as female.

As illustrated in Fig. 1a, we assembled a dataset of papers published between 2020 and 2024[1] that included a CDS and compared them with matched papers from the same journals that did not. After matching on publication characteristics and excluding papers without reference records, the final sample includes 216 papers with CDSs and 432 matched papers (total N=648 papers), together citing 40,719 references. For each cited author across these 648 papers, we inferred gender and cultural origin from the name, allowing us to examine whether papers

[1] The dataset was collected in May 2024.

adopting CDSs differ from their matched comparison papers not only in how often they cite women, but also in which women benefit from those shifts in citation practice. We estimate a series of paper-level regressions across matched papers where the outcomes of interest are various measures of citation composition and the primary predictor is always an indicator for the inclusion of a CDS, and where we control for publication year, field, region, team size, and citation impact (see SI Section S2 for details).

We find that papers with citation diversity statements cite more women than matched papers without them. The share of women's names cited is 7.0 percentage points higher in papers with CDSs than in matched comparisons (30.8% vs. 23.8%; Fig. 1b), and this difference remains substantial after controlling for publication year, field, region, team size, and citation impact ($\beta = 0.0709$, $p < 0.001$; Table S2). However, that gain is not evenly distributed. The same papers show almost no differences in citations to men's names ($\beta = 0.0108$, $p = 0.297$), but a much sharper gap in citations to gender-blind names ($\beta = -0.0860$, $p < 0.001$). This suggests the apparent improvement does not reflect a simple redistribution from men to women, but a shift away from names whose gender is less legible to the system.

This asymmetry becomes sharper when we examine the cultural origin of cited names. Papers with CDSs cite significantly more authors with English-origin names ($\beta = 0.0606$, $p < 0.001$), but fewer authors with Chinese names ($\beta = -0.0316$, $p < 0.001$), Japanese names ($\beta = -0.0101$, $p < 0.001$), and Arab names ($\beta = -0.0080$, $p < 0.001$), among others (Fig. 1c, Table S3). In other words, although CDSs are associated with increased citation of women overall, those gains accrue disproportionately to women whose names remain legible as female to English-oriented inference systems. And people whose names lose gender cues in transliteration do not share equally in the benefits of the intervention.

Overall, these results reveal a paradox at the heart of name-based equity interventions. A practice intended to make citation behavior more gender equitable is associated with improvements in recognition for women overall, but it is also associated with more recognition for women whose names are easiest to classify as female. Because this evidence comes from observational data, however, it cannot by itself establish that the intervention causes this asymmetry, nor can it

distinguish linguistic legibility from a number of alternative explanations for the detected relationships. We therefore turn next to preregistered experiments that isolate the causal effect of gender-equity interventions and test whether the same asymmetry emerges under controlled conditions.

3. Study 2: A causal test of the legibility gap

To isolate the causal effect of equity-promoting interventions, we conducted a preregistered experiment (https://aspredicted.org/py94-n9g5.pdf) in which participants selected scholars for a real Facebook campaign promoting outstanding academics. We tested whether motivating people to increase gender diversity disproportionately benefits women with Western names over those with Eastern names. As illustrated in Fig. 2a, we recruited 750 participants from the United States and asked each to choose scholars from a list containing both men and women from Western and Eastern countries to include in the advertisement. After making an initial set of selections, participants were randomly assigned to one of two conditions. In the control condition, they received feedback summarizing non-gender attributes of their selections. In the treatment condition, they were instead told what percentage of their selected scholars were women, following an established design for boosting the selection of women [33]. All participants then chose one additional scholar, allowing us to test whether gender feedback changed who benefited from the intervention (see SI Section S3 for full stimuli and design details). The key question was whether feedback designed to increase gender diversity would benefit all women equally, or primarily those whose names remain clearly legible as female in English.

Following our pre-registered plan, we find that gender feedback increased the selection of women, consistent with prior studies [33]. Yet importantly, here we find that this gender feedback did not benefit all women equally. Participants randomly assigned to receive gender feedback were more likely to choose a Western woman as their final pick than participants in the control condition (21.1% versus 10.8%; $\beta = 0.1029, p < 0.001$; Fig. 2b, Table 1 Model 1). By contrast, the same feedback had no statistically significant effect on the likelihood of selecting an Eastern woman (10.8% versus 8.2%; $\beta = 0.0265, p = 0.217$; Fig. 2b, Table 1 Model 3). This difference remains substantial after controlling for participants' demographics (Table 1 Model 2

and 4). A pre-registered Wald test comparing these treatment effects controlling for demographics confirms that the intervention benefited Western women significantly more than Eastern women ($\chi^2(1) = 4.450$, $p = 0.0349$).

Study 2 thus provides causal evidence for the pattern documented by Study 1. An intervention designed to increase women's representation does increase women's selection overall, but the benefit is concentrated among women whose names more clearly signal female identity in English. One possible explanation, however, is that this asymmetry reflects the cultural background of the evaluators themselves: perhaps U.S. participants respond more readily to Western names because those names are more familiar to them. We therefore turn to a second preregistered experiment designed to distinguish linguistic legibility from cultural familiarity.

4. Study 3: Testing whether shared cultural background closes the legibility gap

If the legibility gap primarily reflects cultural unfamiliarity, then it should weaken when evaluators share the relevant naming context. To test this possibility, our second preregistered experiment recruited participants from both Western and Eastern countries and tested whether evaluators were more responsive to the intervention when judging women from their own broader cultural background. This design allows us to distinguish two mechanisms: a familiarity account, in which shared cultural context should reduce the asymmetry, and a legibility account, in which the asymmetry should persist even across evaluator groups. This study was pre-registered on OSF (https://osf.io/j2ksg).

Here, we recruited 1,500 participants from China, South Korea, Italy, and Germany (375 from each) for a task that closely paralleled Study 2. Each participant selected scholars for a Facebook campaign from a pool containing candidates from all four countries and then received either feedback about the gender distribution of their initial selectees or feedback about other features of their initial selectees before making their final selection. As in Study 2, the primary outcome was whether the final selected scholar was a woman from a Western or Eastern country (see SI Section S4.1 for details).

We first replicated the basic asymmetry from Study 2. Among 1,443 participants who completed the task, gender feedback substantially increased the selection of Western women, from 7.8% in the control condition to 21.4% in the treatment condition ($\beta = 0.1360, p < 0.001$; Table 2 Model 1). By contrast, it produced no statistically significant increase in the selection of Eastern women ($\beta = 0.0184, p = 0.233$; Table 2 Model 3). A pre-registered Wald test comparing these two coefficients confirms that the difference in treatment effects is statistically significant ($\chi^2(1) = 21.157, p < 0.001$). Thus, even in a more culturally heterogeneous sample, the intervention again benefited Western women far more than Eastern women.

Crucially, this pattern held even when evaluators shared their broad cultural background with the women they were evaluating. Among Western participants, gender feedback increased the selection of Western women from 7.8% to 25.3%; among Eastern participants, it increased the selection of Western women from 7.8% to 17.7%. However, the same feedback had little effect on selections of Eastern women by either participant group: among Western participants, selection rose only from 9.2% to 11.5%, and among Eastern participants, it only rose from 8.1% to 9.0% (Fig. 3). An interaction analysis shows that Western participants responded somewhat more strongly than Eastern participants to gender feedback when selecting Western women ($\beta_{Western_woman:\, feedback \times western_participant} = 0.0750, p = 0.041$; Table 2 Model 2), but no comparable difference between participant groups emerged when selecting Eastern women ($\beta_{Eastern_woman:feedback \times western_participant} = 0.0188, p = 0.543$; Table 2 Model 4). A pre-registered Wald test comparing these two interaction terms finds they do not differ significantly ($\chi^2(1) = 1.204, p = 0.272$), providing no evidence that participant region differentially moderates the feedback effect on Western- versus Eastern-woman selection. In follow-up analyses where we examined the decisions by Chinese, South Korean, Italian and German participants about women from their own culture, separately, we find exactly the same pattern of results (SI Section S4.2 and Table S5).

Study 3 therefore sharpens the mechanism underlying the legibility gap. The asymmetry is not simply a byproduct of Western evaluators being unfamiliar with Eastern names. Even evaluators from the same Eastern countries conferred substantially more benefit to Western women than to

Eastern women when exposed to a nudge designed to promote gender diversity. Hence, what matters is not merely who is judging, but whether female identity remains legible in the names on which the intervention depends. Together, Studies 2 and 3 show that the unequal gains documented in Study 1 are consistent with benefits that accrue causally: when policies rely on inferred gender, linguistic legibility determines who is recognized as a woman and who benefits from efforts to support women in science.

5. Discussion

Our findings reveal a structural paradox in contemporary efforts to promote gender equity in science. Interventions designed to increase women's recognition do not benefit all women equally when they rely on names to infer gender. Instead, they disproportionately advantage women whose names remain legible as female in English, while bypassing women whose gender cues are obscured in transliteration. Across observational and experimental evidence, we show that progress toward gender parity can therefore come with an unrecognized tradeoff: gains in average inclusion may mask a redistribution of recognition across linguistic and cultural lines.

Name-based inference approaches make efforts to measure and reduce potential gender gaps more scalable and feasible across large datasets and organizations, and they have become a key part of the infrastructure through which institutions measure representation, diagnose inequity, and design corrective interventions. In that context, errors in gender inference are not merely measurement problems [34-36]. They can become allocation problems. When some identities are more legible than others, the very tools designed to improve fairness can channel recognition toward those whose identities are most legible to the system. Our results thus shift attention from asking whether equity interventions work on average to addressing a deeper question: for whom do they work, and under what conditions?

Our experiments clarify the mechanism underlying this asymmetry. The unequal gains we document are not simply the product of Western evaluators favoring culturally familiar names. Even when evaluators come from Eastern countries, interventions intended to increase women's representation continued to benefit Western women far more than Eastern women. What matters,

then, is not just who is judging, but whether female identity remains legible in the names on which judgment depends. In this sense, the legibility gap is interpreter-general: it emerges in both algorithmic and human inference when names function as the proxy through which gender is encoded.

Our findings also illuminate an overlooked tension in the current turn toward data-driven equity [37]. Over the past decade, scholars, journals, and institutions have increasingly embraced inclusive citation practices, equity dashboards, and algorithmic auditing tools to monitor representation and correct disparities. Here, our results show that when inclusion is operationalized through proxies such as names, the universality of measurement can conceal the cultural specificity of the signals being measured. A system may appear neutral because it applies the same rule to everyone, while still systematically favoring those whose identities are more easily legible within the linguistic conventions on which that rule depends [38,39].

The broader implication extends beyond gender classification. In many global evaluative settings, recognition depends on signals that must travel across languages, scripts, and cultural contexts [40-42]. When those signals survive unevenly, systems may reward what remains easiest to interpret [43]. Legibility, in this sense, represents a hidden dimension in evaluation systems. The question is not just who is represented, but whose identity remains legible enough to be counted in the first place. This suggests a broader agenda for studying fairness in transnational and algorithmically mediated settings.

Several limitations qualify our analysis and point to important future directions. Our focus on the distinction between specific “Western” and “Eastern” countries’ naming systems is somewhat coarse, and gender legibility likely varies within countries, across diasporas, and across hybrid naming practices. Our analyses also inherit the binary gender assumptions built into existing inference tools, leaving open important questions about nonbinary identities and culturally specific gender systems [44]. Finally, we focus on immediate outcomes in citation and selection; future work should examine whether legibility-based advantages compound over time through cumulative processes of visibility, collaboration, and advancement.

Taken together, our results highlight a fundamental challenge for equity in an increasingly global and algorithmic scientific system. Inclusion depends not only on whether interventions expand opportunity on average, but also on whether the categories through which opportunity is measured remain legible across languages and cultures. The legibility gap shows how infrastructures designed to enhance fairness can silently reproduce the hierarchies of the languages that feed them. In global systems of recognition, equity depends not only on equal opportunity, but on equal legibility.

Data and Code Availability: All analysis code, data files, and an extended appendix containing additional tables are available on the Open Science Framework at https://osf.io/jgacu/overview?view_only=db6848c8fea645a8bf45f0c090956d56

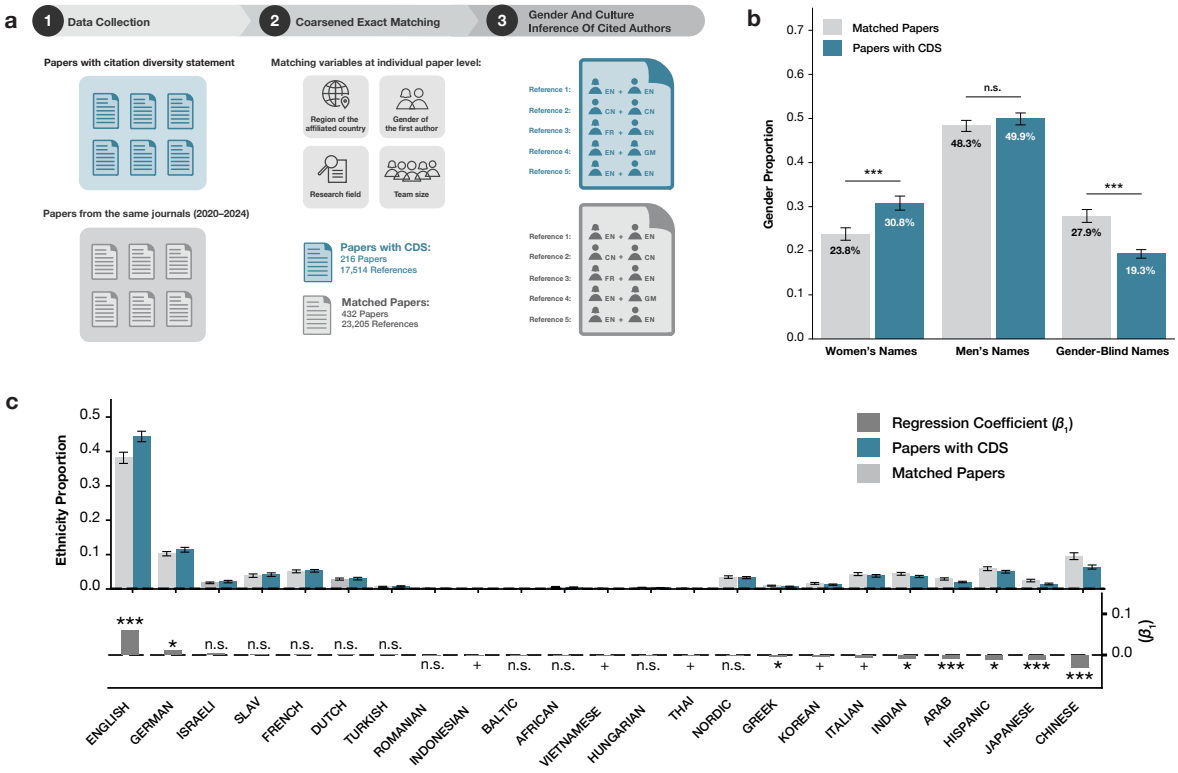


Figure 1. Citation diversity statements and the redistribution of recognition. (a) We identified 239 research articles containing a citation diversity statement and assembled a matched comparison group using Coarsened Exact Matching (CEM), yielding 216 papers with a CDS and 432 matched papers. For each paper, we extracted the first- and last-author names of their references and inferred (1) gender using the *gender-guesser* library and (2) cultural origin using *Ethnea*. (b) Gender composition of cited authors. Papers with citation diversity statements cited women (30.8%) at a rate 7.0 percentage points higher than the matched sample (23.8%). This difference was offset by a significantly lower (8.6 percentage points lower) rate of citations to gender-blind names, and there was no corresponding difference in citations to men, suggesting a preference for names that clearly signal a female identity. (c) Cultural composition of cited authors' names. Papers with citation diversity statements cite significantly more authors with English-origin names, while citing several non-Western cultural groups less frequently. Bars display means ± 1 standard error and are ordered by regression coefficients. $+\ p < 0.10, *\ p < 0.05, **\ p < 0.01, ***\ p < 0.001$.

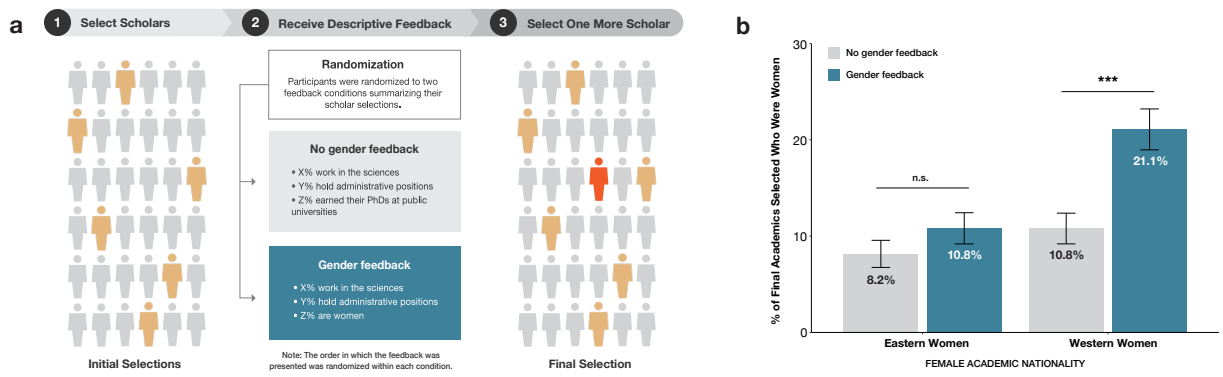


Figure 2. The effect of gender feedback on the selection of a seventh female scholar, split by scholar nationality (Study 2). (a) This panel illustrates the experimental design workflow. Participants were first presented with a list of scholars from which they selected six to feature in a Facebook advertisement about academics. In Step 2, participants were randomly assigned to one of two feedback conditions, each providing a summary of the characteristics of the scholars they had initially selected. In the gender-feedback condition, participants received feedback summarizing the gender composition of their six initial selections; in the no-gender-feedback condition, no gender feedback was provided. In Step 3, participants were asked to select a seventh scholar, and the gender identity of their final selectee is our primary dependent measure. (b) The share of participants who chose a female scholar as their final (seventh) pick, disaggregated by whether the scholar was from an Eastern or Western background and by experimental condition (gender-feedback vs. no-gender-feedback). The x-axis distinguishes scholars from Eastern countries (China, South Korea, Thailand) and Western countries (United Kingdom, Italy, Germany). Grey bars represent the no-gender-feedback condition, in which participants received no feedback summarizing the gender mix of their initial six selections; blue bars represent the gender-feedback condition, in which participants were given feedback summarizing the gender composition of their initial six selections before picking a seventh. Error bars denote ± 1 standard error. $+\ p < 0.1$; $*\ p < 0.05$; $**\ p < 0.01$; $***\ p < 0.001$.

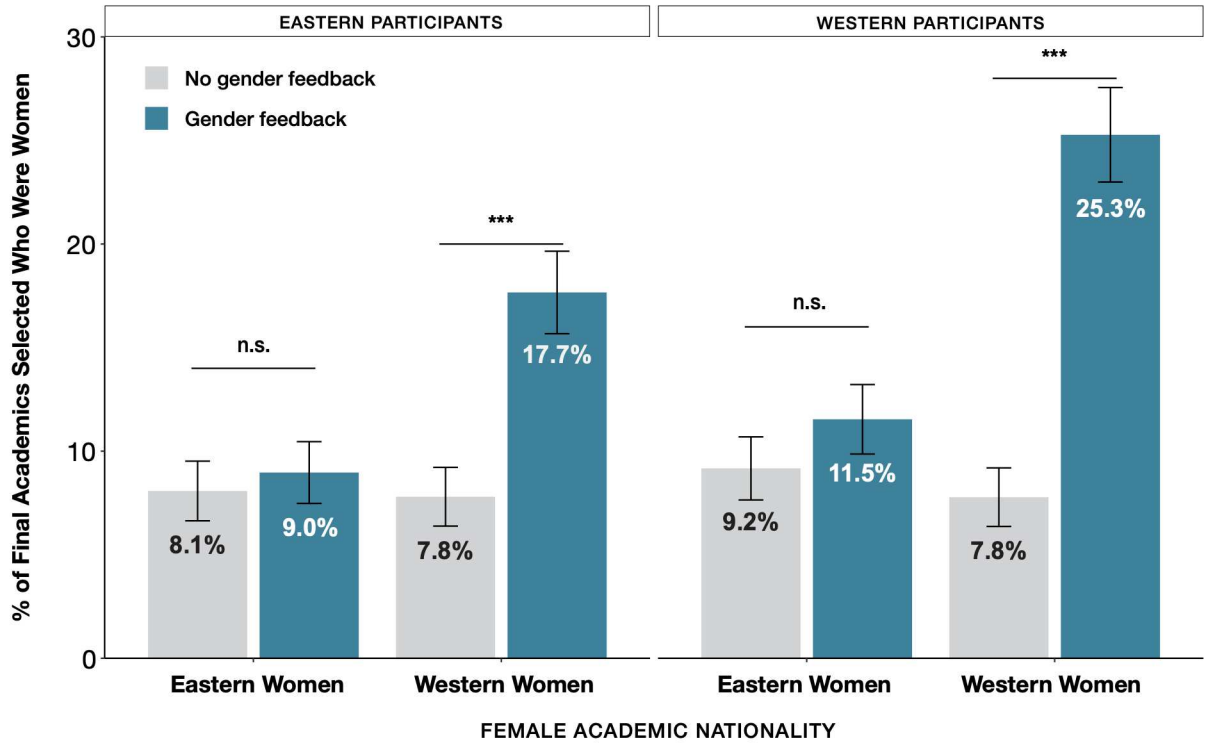


Figure 3. The effect of gender feedback on the selection of a seventh female scholar, split by participant region and scholar nationality (Study 3). The chart shows the share of participants who chose a female scholar as their final (seventh) pick in Study 3, disaggregated by whether the scholar was from an Eastern or Western background and by experimental condition (gender feedback intervention vs. control). The Left panel shows data from participants from Eastern countries (China, South Korea), and the Right panel shows data from participants from Western countries (Germany, Italy). Within each panel, the x-axis distinguishes between selected scholars from Eastern nations (China or South Korea) and Western nations (Germany or Italy). Grey bars represent the *no-gender-feedback* condition, in which participants received no feedback summarizing the gender mix of their initial six selections; blue bars represent the *gender-feedback* condition, in which participants were given feedback summarizing the gender composition of their initial six selections before picking a seventh scholar. Error bars denote ± 1 standard error. $+p < 0.1$; $* \ p < 0.05$; $** \ p < 0.01$; $*** \ p < 0.001$.

Dependent variable: Identity of final academic selected:	Western Woman		Eastern Woman	
	Model 1	Model 2	Model 3	Model 4
Intervention (Feedback was provided on the % of academics who were women)	0.1029***	0.1037***	0.0265	0.0268
	(0.0266)	(0.0267)	(0.0215)	(0.0216)
Controls	No	Yes	No	Yes
Observations	750	750	750	750
R-Squared	0.0198	0.0259	0.0021	0.0030
Wald Test (Intervention in Model 2 vs. Model 4)	$\chi^2(1) = 4.450$, $p = 0.0349$			

Table 1. OLS Regressions estimating the effect of gender feedback on participants' final selection (Study 2). The table presents four OLS regressions predicting whether the seventh (final) scholar chosen by a participant for inclusion in a Facebook advertisement was a Western woman (Models 1–2) or an Eastern woman (Models 3–4). The key independent variable in every model is a binary indicator for random assignment to receive feedback on the percentage of women in the participant's initial six-scholar roster. Models 2 and 4 add demographic controls: participant gender (male = 1), race (White = 1), and age. The Wald test row reports a cross-equation test comparing the Intervention coefficient across Model 2 and Model 4. Robust (HC3) standard errors are shown in parentheses. $+p < 0.1$; $* p < 0.05$; $** p < 0.01$; $*** p < 0.001$.

Dependent variable: Identity of final academic selected:	Western Woman		Eastern Woman	
	Model 1	Model 2	Model 3	Model 4
Intervention (Feedback was provided on the % of academics who were women)	0.1360*** (0.0183)	0.0985*** (0.0259)	0.0184 (0.0154)	0.0090 (0.0218)
Intervention x Western Participant		0.0750* (0.0366)		0.0188 (0.0308)
Western Participant	0.0066 (0.0341)	-0.0323 (0.0390)	0.0303 (0.0287)	0.0205 (0.0329)
Controls	Yes	Yes	Yes	Yes
Observations	1,443	1,443	1,443	1,443
R-Squared	0.0413	0.0441	0.0034	0.0036
Wald Test (Intervention in Model 1 vs. Model 3)	$\chi^2(1) = 21.157, p < 0.001$			
Wald Test (Intervention x Western Participant in Model 2 vs. Model 4)	$\chi^2(1) = 1.204, p = 0.272$			

Table 2. **OLS Regressions estimating the effect of gender feedback on participants' final selection (Study 3).** The table presents four OLS regressions predicting whether the seventh (final) scholar chosen by a participant for inclusion in a Facebook advertisement was a Western woman (Models 1–2) or an Eastern woman (Models 3–4). The key independent variable in every model is a binary indicator for random assignment to receive feedback on the percentage of women in the participant's initial six-scholar roster. Models 2 and 4 add an interaction with the participant's region of residence (Western = Germany/Italy; Eastern = China/South Korea). All models include demographic controls and incentive fixed effects: participant gender (male = 1), race (White = 1), and age. After excluding participants who did not complete the experiment, the analytic sample comprises 1,443 of the 1,500 recruited participants. The Wald test rows report cross-equation tests comparing the Intervention coefficient across Model 1 and Model 3, and the Intervention x Western Participant coefficient across Model 2 and Model 4. Robust (HC3) standard errors appear in parentheses. $+\ p < 0.1$; $*\ p < 0.05$; $**\ p < 0.01$; $***\ p < 0.001$.

Reference:

1 Etzkowitz, H., Kemelgor, C., Neuschatz, M., Uzzi, B. & Alonzo, J. The paradox of critical mass for women in science. *Science* **266**, 51–54 (1994).

2 Ding, W. W., Murray, F. & Stuart, T. E. Gender differences in patenting in the academic life sciences. *Science* **313**, 665–667 (2006).

3 Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, J. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* **109**, 16474–16479 (2012).

4 Larivière, V., Ni, C., Gingras, Y., Cronin, B. & Sugimoto, C. R. Bibliometrics: Global gender disparities in science. *Nature* **504**, 211–213 (2013).

5 Hofstra, B. *et al.* The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences* **117**, 9284–9291 (2020).

6 Huang, J., Gates, A. J., Sinatra, R. & Barabási, A.-L. Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences* **117**, 4609–4616 (2020).

7 Myers, K. R. *et al.* Unequal effects of the COVID-19 pandemic on scientists. *Nature Human Behaviour* **4**, 880–883 (2020).

8 Dworkin, J. D. *et al.* The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience* **23**, 918–926 (2020).

9 Koning, R., Samila, S. & Ferguson, J.-P. Who do we invent for? Patents by women focus more on women's health, but few women get to invent. *Science* **372**, 1345–1348 (2021).

10 Morgan, A. C. *et al.* The unequal impact of parenthood in academia. *Science Advances* **7**, eabd1996 (2021).

11 Ni, C., Smith, E., Yuan, H., Larivière, V. & Sugimoto, C. R. The gendered nature of authorship. *Science Advances* **7**, eabe4639 (2021).

12 Vásárhelyi, O., Zakhlebin, I., Milojević, S. & Horvát, E.-Á. Gender inequities in the online dissemination of scholars' work. *Proceedings of the National Academy of Sciences* **118**, e2102945118 (2021).

13 Ross, M. B. *et al.* Women are credited less in science than men. *Nature* **608**, 135–145 (2022).

14 Yang, Y., Tian, T. Y., Woodruff, T. K., Jones, B. F. & Uzzi, B. Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proceedings of the National Academy of Sciences* **119**, e2200841119 (2022).

15 Spoon, K. *et al.* Gender and retention patterns among US faculty. *Science Advances* **9**, eadi2205 (2023).

16 Peng, H., Teplitskiy, M., Romero, D. M. & Horvát, E.-Á. The gender gap in scholarly self-promotion on social media. *Nature Communications* **16**, 5552 (2025).

17 Sowrirajan, T., Whalen, R. & Uzzi, B. The institutional dynamics of inequality for women inventors who break with conventional thinking. *Proceedings of the National Academy of Sciences* **123**, e2500343123 (2026).

18 Vishkin, A., Slepian, M. L. & Galinsky, A. D. The gender-equality paradox and optimal distinctiveness: more gender-equal societies have more gendered names. *Social Psychological and Personality Science* **13**, 490–499 (2022).

19 Zwebner, Y., Miller, M., Grobgeld, N., Goldenberg, J. & Mayo, R. Can names shape facial appearance? *Proceedings of the National Academy of Sciences* **121**, e2405334121 (2024).
20 Allen, R. E. & Wiles, J. L. A rose by any other name: Participants choosing research pseudonyms. *Qualitative Research in Psychology* **13**, 149–165 (2016).
21 Kirgios, E. L., Rai, A., Chang, E. H. & Milkman, K. L. When seeking help, women and racial/ethnic minorities benefit from explicitly stating their identity. *Nature Human Behaviour* **6**, 383–391 (2022).
22 Kline, P., Rose, E. K. & Walters, C. R. Systemic discrimination among large US employers. *The Quarterly Journal of Economics* **137**, 1963–2036 (2022).
23 Bertrand, M. & Mullainathan, S. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* **94**, 991–1013 (2004).
24 U.S. Equal Employment Opportunity Commission. *Pre-Employment Inquiries and Gender*, 2026).
25 Lockhart, J. W., King, M. M. & Munsch, C. Name-based demographic inference and the unequal distribution of misrecognition. *Nature Human Behaviour* **7**, 1084–1095 (2023).
26 Santamaría, L. & Mihaljević, H. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science* **4**, e156 (2018).
27 VanHelene, A. D. *et al.* Inferring gender from first names: Comparing the accuracy of Genderize, Gender API, and the gender R package on authors of diverse nationality. *PLOS Digital Health* **3**, e0000456 (2024).
28 Montgomery, S. L. *Does science need a global language?: English and the future of research*. (University of Chicago Press, 2013).
29 Boussebaa, M. & Tienari, J. Englishization and the politics of knowledge production in management studies. *Journal of Management Inquiry* **30**, 59–67 (2021).
30 Lin, Z., Yin, Y., Liu, L. & Wang, D. SciSciNet: A large-scale open data lake for the science of science research. *Scientific Data* **10**, 315 (2023).
31 Zurn, P., Bassett, D. S. & Rust, N. C. The citation diversity statement: a practice of transparency, a way of life. *Trends in Cognitive Sciences* **24**, 669–672 (2020).
32 Citation diversity statements. *Nature Reviews Psychology* **4**, 617–617 (2025).
33 Cervantez, J. A., Pink, S. L., Chang, L., Rai, A. & Milkman, K. L. Does feedback enhance diversity in selection decisions. *Management Science*, (in press) (2026).
34 Jacobs, A. Z. & Wallach, H. in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency.* 375–385.
35 Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
36 Kleinberg, J., Ludwig, J., Mullainathan, S. & Rambachan, A. Algorithmic Fairness. *AEA Papers and Proceedings* **108**, 22–27 (2018).
37 Espeland, W. N. & Stevens, M. L. A sociology of quantification. *European Journal of Sociology/Archives Européennes de Sociologie* **49**, 401–436 (2008).
38 Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S. & Vertesi, J. in *Proceedings of the Conference on Fairness, Accountability, and Transparency.* 59–68.
39 Rivera, L. A. & Tilcsik, A. Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation. *American Sociological Review* **84**, 248–274 (2019).

40 Quillian, L. & Lee, J. J. Trends in racial and ethnic discrimination in hiring in six Western countries. *Proceedings of the National Academy of Sciences* **120**, e2212875120 (2023).
41 Falk, A. & Hermle, J. Relationship of gender differences in preferences to economic development and gender equality. *Science* **362**, eaas9899 (2018).
42 Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behavioral and Brain Sciences* **33**, 61–83 (2010).
43 Rivera, L. A. Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review* **77**, 999–1022 (2012).
44 Keyes, O. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction* **2**, 1–22 (2018).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- WangSIv3.pdf
- WangSIv4.pdf