First authorship gender gap in the geosciences

Tamara Pico[1], Paul Bierman[2], Sarah Richardson[3], and Kevin Doyle[4]

[1]Caltech
[2]University of Vermont
[3]Harvard University
[4]Diffeo

November 30, 2022

Abstract

Although gender parity has been reached at the graduate level in the geosciences, women remain a minority in top-level positions. First authorship of peer-reviewed scholarship is a measure of academic success and is often used to project potential in the hiring process.

Given the importance of first author publications for hiring and advancement, we sought to quantify whether women are underrepresented as first authors relative to their representation in the field. We compiled first author names across 13 leading geoscience journals from January 2013 to April 2019 (n = 35,183). Using a database of 216,286 names from 79 countries, across 89 languages, we classified the likely gender associated with each author's given (first) name. We also estimated the gender distribution of authors who publish using only initials, which may itself be a strategy employed by some women to preempt perceived (and actual) gender bias in the publication process. Female-author names represent 13-30% of all first authors in our database, and are significantly underrepresented relative to the proportion of women in early career positions (30-50%). The proportion of female-name first authors varies significantly by subfield, reflecting variation in representation of women across subdisciplines. In geoscience, the quantification of this first authorship gender gap supports the hypothesis that the publication process – namely, achievement or allocation of first authorship – is biased by social factors, which may modulate career success of women in the sciences.

First authorship gender gap in the geosciences

T. Pico[1]*, P. Bierman[2], K. Doyle[3], S. Richardson[4]

* Corresponding author: tpico@g.harvard.edu

[1] Division of Geological and Planetary Sciences, Caltech, Pasadena, CA

[2] Department of Geology, University of Vermont, Burlington, VT

[3] Diffeo, Cambridge, MA

[4] Department of History of Science and Studies of Women, Gender, and Sexuality, Harvard University, Cambridge, MA

Keywords: science careers; gender bias; scholarly productivity; geoscience

Abstract

Although gender parity has been reached at the graduate level in the geosciences, women remain a minority in top-level positions. First authorship of peer-reviewed scholarship is a measure of academic success and is often used to project potential in the hiring process. Given the importance of first author publications for hiring and advancement, we sought to quantify whether women are underrepresented as first authors relative to their representation in the field. We compiled first author names across 13 leading geoscience journals from January 2013 to April 2019 (n = 35,183). Using a database of 216,286 names from 79 countries, across 89 languages, we classified the likely gender associated with each author's given (first) name. We also estimated the gender distribution of authors who publish using only initials, which may itself be a strategy employed by some women to preempt perceived (and actual) gender bias in the publication process. Female-author names represent 13-30% of all first authors in our database, and are significantly underrepresented relative to the proportion of women in early

career positions (30-50%). The proportion of female-name first authors varies significantly by subfield, reflecting variation in representation of women across subdisciplines. In geoscience, the quantification of this first authorship gender gap supports the hypothesis that the publication process – namely, achievement or allocation of first authorship – is biased by social factors, which may modulate career success of women in the sciences.

Introduction

First authorship of papers in peer-reviewed journals is crucial to academic success, promotion, and competitive research funding (1,2). Authorship is key to moving up the career ladder from graduate school to postdoctoral positions to faculty appointments (3). In the natural sciences, women are underrepresented at the highest academic tiers (4–7). Representation of women in academic geoscience drops off significantly at every successive tier, with the greatest discrepancy at the highest ranks. This representation varies by career stage and subfield (40-50% of Ocean, Atmospheric, and Earth Sciences graduate students (8), 30-36% at the assistant professor level, and only 11.5-13% at the full professor level (6,7) are women).

A critical contributor to this gender gap is the transition from post-doc to the first faculty position (9), and studies suggest this discrepancy results from differences in academic productivity and perceived potential (3). While academic productivity, measured by publication record, is often assumed to represent inherent scientific talent (10), the strongest predictor of scholarly productivity is work environment, which highlights the importance of social factors in

46 determining academic success (11). For decades, publication analyses have revealed a
47 significant gender gap in authorship (12), publication in high impact journals (13), and citation
48 rates (14,15). While recent assessments document the persistence of a gender discrepancy in
49 first authorship of peer-reviewed publications in the sciences (4,5,16–18), an in-depth study
50 focused on the geosciences has yet to be done. Analysis of authorship imbalances contributes
51 to a stream of recent scholarship quantifying gender inequities in the geosciences at research
52 conferences (19,20), in peer review (21), and in recommendation letters (9).
53
54 Given the importance of first authorship for career advancement (3), we sought to assess the
55 extent to which female first authors are underrepresented among 13 of the major geoscience
56 journals. In this field, it is first authors who conventionally perform the majority of the research
57 and the writing. We used data-mining to quantify the representation of women as first authors
58 from January 2013 to April 2019 in leading geoscience journals (*Nature Geoscience, Geology,*
59 *Geological Society of America Bulletin, Journal of Geophysical Research (JGR) – all fields,*
60 *Geophysical Research Letters, Quaternary Science Reviews, Geochimica et Cosmochimica*
61 *Acta)*. 62% of first-author names were categorizable by gender (Table 1).
62
63 One factor that potentially confounds any analysis of women's representation in science is that
64 women may be more likely to initialize their given name in order to mask their gender as a
65 preemptive defense against implicit bias (as substantiated by studies showing that a name's
66 gender influences competence assessments (22)). In this study, we compared initialed author
67 names to all authors in the complete mined database and identified the likely given name

based on coauthorship overlap. We then assigned a likely gender to these first authors with initialized given names in order to assess the extent to which the practice of initializing names impacts measures of women's representation in the geoscience. We include open-access code in a GitHub repository to reproduce this approach in future studies, because quantifying authorship gender ratios will be useful to repeat for specific subdisciplines as well as to test for change over time (see Materials & Methods).

Results

In the majority of journals analyzed (ten of the thirteen), female names made up fewer than 30% of gender-categorizable first-author names. The proportion of female name first authors varies significantly by subfield, and likely reflects the representation of women across subdisciplines. We found that female names represented at lowest, 23% of categorizable first author names in *Journal of Geophysical Research Space Physics (*where representation of women in student or early career positions is close to 20% (23)), and at most 36% of categorizable first author names in *Journal of Geophysical Research Biogeosciences* (n =26,623 for categorizable first author names, excluding unmatched initials; see Materials & Methods) across the journals analyzed (Figure 1; Figure 2).

Of the full database, including non-categorizable names, the percentage of female names ranges from 13-30% across all journals. Male names (green; Figure 1) represent 25-61% of all names, while uncategorized names (black; Figure 1) and unmatched initials (purple; Figure 1)

represent 5-16% and 11-48% of all names, respectively. Early career scientists, defined as those who received their highest degree within the last 10 years, constitute the majority (~70%; see Methods) of first authors in geoscience; however, the percentage of female names (13-30%) is significantly below the representation of women at this career level (30-50%; translucent purple bars; Figure 1; Figure 2).

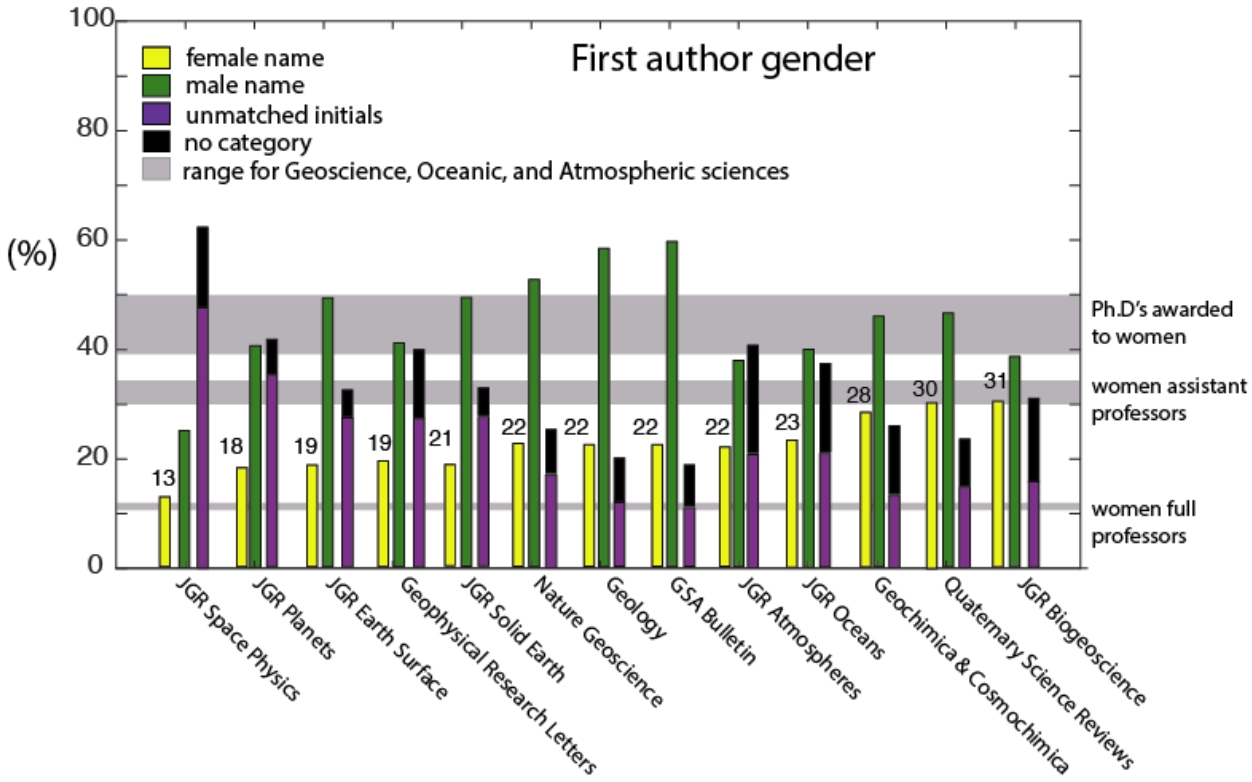

Figure 1 | Representation (%) of female names (13-30%; yellow), male names (25-61%; green), unmatched initials (11-48%; purple), and uncategorized (5-16%; black) non-initialed names in total first authors across geoscience journals between January 2013 and April 2019. Female names (%) labeled for each journal. Light purple bars show representation of Ph.D's awarded to women in 2016 (8), women assistant professors and women full professors (6,7) in 2010 in Geosciences, Oceanic, and Atmospheric sciences.

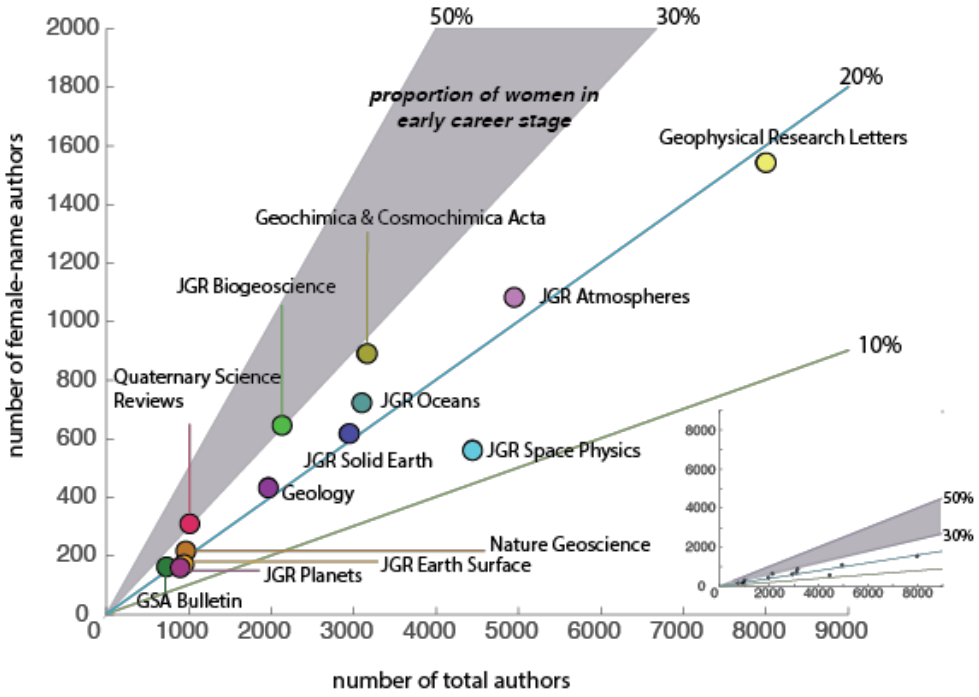

Figure 2 | Proportion of female-name authors for each journal. Inset figure shows equal axes for number of total authors and number of female-name authors. The purple shaded region spans 30-50%, the proportion range of women in early career positions in geoscience (from women in assistant professor positions to women earning Ph.D's). The green and blue lines represent the 10% and 20% proportion line, respectively.

Of the matched initialed given names, we found that 29% were categorizable as female names (n =417, out of 1,434). This percentage varied by journal from 14% (*Geological Society of America Bulletin*) to 41% (*Quaternary Science Reviews*). In *Geological Society of America Bulletin*, female names represent 25% of all first authors, suggesting that men are more likely to publish using initials in this journal, whereas in *Quaternary Science Reviews* female names are slightly over-represented in initialed names (female names represent 36% of all first authors). Although we are unable to match all initialed first author names, the percentage of female names in matched initialed given names (29%) is proportional to the overall representation of

119 female names across all 13 journals (28%), indicating there is not a significant gender bias in
120 authors' decision to publish in the geosciences using only initials.
121
122 **Discussion**
123 Geoscience is not the only field with a first author gender gap. In other disciplines, a similar first
124 authorship gender gap was quantified by Bendels et al., including in the biological sciences
125 (female names represent 35% of first authors) and chemistry (female names represent 23% of
126 first authors); this gap persists across communities internationally (4). A study analyzing
127 neuroscience journals showed that female name first authorship only increased by 1% from
128 2006 to 2016 (17). Bendels et al. found a comparable first authorship gender gap in the Earth &
129 Environmental sciences (24%), and reported an annual growth rate of 1.8% for female name
130 first authorship (4). If we assume this growth rate of female name first authorship continues,
131 we estimate that parity would be reached in Earth & Environmental sciences in 2061.
132
133 The results from this study are limited by the range of journals selected for analysis and the
134 specific subfields of geoscience these journals represent. Future studies could reproduce this
135 analysis with other subdiscipline-specific journals using the open-access code provided from
136 this study in a GitHub repository (see Materials & Methods). One limitation to our approach is
137 the choice of gender-categorizing method and database. For example, genderize.io will not be
138 able to identify the gender for names pertaining to cultures where given names are not
139 gendered (e.g. some East Asian cultures). Furthermore, in this study, we compared our results
140 to the representation of women in geoscience within the United States, even though the author

141 names included in this study come from a range of international institutions, and the
142 proportion of women in different geoscience career stages varies by country.
143
144 We cannot draw a firm conclusion about what drives the identified disparity in first authorship
145 but we can speculate based on the existing literature. Biases may exist at many different stages
146 of the publication process. At the graduate school level, women may receive less mentoring or
147 encouragement to write and submit first-author research articles (24,25). A study analyzing
148 authorship in political science journals found a gender bias in the perception of likely
149 acceptance in journals, and therefore, in the ultimate decision to submit articles (26).
150
151 Double blind review, which is not widely used in the geosciences, has been shown to reduce
152 gender gaps in publication acceptance rates (27), although a study on peer review in ecology
153 suggests that reviewers do not rate papers differently based on first author gender in that field
154 (28). First authors may respond differently to a paper's rejection, as studies on confidence
155 suggest that men's self-assessment of competence is significantly higher than those of women
156 (29,30). Because of this higher level of confidence, men may be more likely to resubmit a paper
157 following a rejection, contributing to a higher rate of male first authorship in top journals. To
158 understand what causes our finding of a gender disparity in first author publication rates, it
159 would be helpful to understand disparities at different stages in the publication process in the
160 geosciences. Are women submitting fewer papers, are women's papers being rejected at higher
161 rates, or do women resubmit at lower rates compared to male counterparts? Answering these

questions would require journals to track gender (in addition to other social metrics such as career stage) in submitted and accepted manuscripts.

As with gender, journals could consider other demographic sources of inequities such as race. However, it is more tractable to infer the gender of given names than to identify race. For many journals, first author demographics are not tracked at submission, and therefore self-reported gender or race data are not available. Improved datasets documenting representation by gender, race/ethnicity, sexuality, and nationality, across different career stages in a range of disciplines, may help identify where biases exist in the publication pipeline (31).

Our findings support efforts to implement journal practices, such as double-blind review, which reduce the impact of perceptions of first author gender and have been shown to increase the success of women in publishing articles (27). In addition, mentoring is an important element in academic productivity for early career scientists, and gender has been shown to influence the degree of mentorship provided (25). The gender-pairing of faculty mentors with students can result in different scholarly productivities (32), and links between gender, mentoring, and publication might be highlighted by institutional leaders to raise awareness around social bias in mentoring. Scientific communities might also consider other ways to recognize the various contributions of authors, reevaluating the weight placed on first author publications (33).

Data documenting gender biases in the publication process in addition to studies identifying the impact of social factors on productivity (1,2,26,30) challenge the view that science careers

184 advance solely on merit. Underrepresentation of female first authors relative to their presence
185 in the geosciences contributes to a growing body of evidence that suggests success in science is
186 strongly modulated by social factors (2), and that these factors influence tangible products such
187 as first-authored publications. Efforts by journals, funders, and professional societies to
188 understand what practices produce gender disparities in scholarly achievement will be required
189 to reduce bias in and out of the publication pipeline.
190
191 **Materials & methods**
192 The code used to produce the results included in this study can be found at
193 https://github.com/kevindoyle/geoscience-first-authorship. We compiled author names from
194 January 2013 to April 2019 across a range of 13 geoscience journals (*Nature Geoscience,*
195 *Geology, Geological Society of America Bulletin, Journal of Geophysical Research (JGR)– all*
196 *fields, Geophysical Research Letters, Quaternary Science Reviews, Geochimica et Cosmochimica*
197 *Acta)*. We selected these journals to include a range of general geoscience journals as well as
198 discipline-specific journals across Earth, Ocean, and Atmospheric Sciences. We web-scraped
199 author and article names from each journal website (n = 35,183) by iteratively changing query
200 parameters in the websites' search page URLs. The search result pages were rendered and
201 downloaded using the Python package selenium (34). Author names and article titles were
202 parsed from the downloaded pages by navigating the HTML tree using the python package
203 BeautifulSoup (35). An author's given name was identified as the first token of an author's
204 name string. Tokens were created using whitespace as a delimiter. Of these author names,
205 24,525 are unique full names and 7,157 are unique given names. We classified the gender of

author's given names using the genderize.io API (36), accessed through a python client (37). The genderize.io database contains 216,286 distinct given names across 79 countries and 89 languages. This library categorizes names as 'female', 'male', or 'uncategorized', and returns the probability that the given name is classified as a specific gender. In running the scraped author names through this database, we assigned the category 'female' or 'male' if the probability was above 50% for the given gender. This approach is limited to gendered given names, which may not hold across all cultures. Furthermore, this approach assumes the first name is the given name, which is not true for some cultures where family names are the first name.

Of 35,183 first author names, 9,994 names were initials (28%). To improve the accuracy of our results, we attempted to identify the non-initialed given name of initialed authors. This was done by comparing initialed names to all authors in the complete database of publications across these 13 journals. For a given initialed name, we used the associated family name (identified as the last token in the name string) to find all articles that included a coauthor with that family name. We then compared the extent of overlap in coauthor names between the list of articles containing this family name. The article with the greatest overlap in coauthorship (minimum overlap of one) was selected to identify the given name of the initialed first author. We were able to match 1,434 of 9,994 (14.3%) of initialed first authors. In calculating the overall representation of female or male names, we combined the matched given names of initialed authors with the remaining set of first authors who published using their non-initialed

given name (a total of 26,623 names). We then identified the likely gender of all compiled given names.

We then compared these percentages to the representation of women at different career stages in geoscience in the United States (translucent purple bars; Figure 1). Because publications do not identify first-author career position (i.e., student, postdoc, faculty), we could not categorize all first author names by likely career stage. To estimate the representation of early career scientists in first author positions in geoscience journals, we used a database of first-author names in American Geophysical Union (AGU) journals from 2013-2018, which have been categorized for career stage. These stages are defined as student, early career, mid-career, experienced, and retired. Here, early career stage is defined as those who received their highest degree within the last 10 years or those within the age range of 30-39 if no graduation date was provided. According to this database, 68% of first authors in AGU journals from 2013-2018 were students or early career stage.

Acknowledgments

We are grateful to K. Blaufuss and B. Hanson for providing a dataset on first authorship demographics in AGU journals. T. Pico acknowledges funding from Harvard University and NSF-GRFP. The code used in this study is available in the following Git Hub repository: https://github.com/kevindoyle/geoscience-first-authorship.

References

1. Evans HK, Houston S. Reflecting on a Decade of Women's Publications in Four Top

Political Science Journals. Am Polit Sci Assoc. 2011;44(4):793–8.

2. Way SF, Morgan AC, Larremore DB, Clauset A. Productivity, prominence, and the effects of academic environment. Proc Natl Acad Sci. 2019;201817431.

3. Lerchenmueller MJ, Sorenson O. The gender gap in early career transitions in the life sciences. Res Policy [Internet]. 2018;47(6):1007–17. Available from: https://doi.org/10.1016/j.respol.2018.02.009

4. Bendels MHK, Brueggmann D, Groneberg DA, Mu R. Gender disparities in high-quality research revealed by Nature Index journals. PLoS One. 2018;1–21.

5. Filardo G, Graca B, Sass DM, Pollock BD, Smith EB, Martinez MA. Trends and comparison of female first authorship in high impact medical journals : observational study (1994-2014). BMJ. 2014;

6. Glass JB. We are the 20%: Updated Statistics on Female Faculty in Earth Sciences in the U.S. Women Geosci Pract Posit Pract Towar Parit. 2015;(May 2015):17–22.

7. Macphee D, Canetto SS. Women in academic atmospheric sciences. Bull Am Meteorol Soc. 2015;96(1):59–67.

8. Bernard RE, Cooperdock EHG. No progress on diversity in 40 years. Nat Geosci [Internet]. 2018;(April). Available from: http://dx.doi.org/10.1038/s41561-018-0116-6

9. Dutt K, Pfa DL, Bernstein AF, Dillard JS, Block CJ. Gender differences in recommendation letters for postdoctoral fellowships in geoscience. Nat Geosci. 2016;9(October).

10. Heesen R. Academic superstars: competent or lucky? Synthese. 2017;194(11):4499–518.

11. Way SF, Morgan AC, Larremore DB, Clauset A. Productivity , prominence , and the effects of academic environment. PNAS. 2019;1–5.

12. Cole JR, Zuckerman H. The Productivity Puzzle: Persistence and change in patterns of publication of men and women scientists. Women Sci. 1984;2.
13. Brooks C, Fenton EM, Walker JT. Gender and the evaluation of research. Res Policy. 2014;43(6):990–1001.
14. Caplar N, Tacchella S, Birrer S. Quantitative evaluation of gender bias in astronomical publications from citation counts. Nat Astron. 2017;0141(May).
15. King MM, Bergstrom CT, Correll SJ, Jacquet J, West JD. Men Set Their Own Cites High : Gender and Self-citation across Fields and over Time. SOCIUS. 2017;3:1–22.
16. West JD, Jacquet J, King MM, Correll SJ, Bergstrom CT. The Role of Gender in Scholarly Authorship. 2013;8(7).
17. Shen YA, Webster JM, Shoda Y, Fine I. Persistent Underrepresenation of Women's Science in High-profile Journals. bioRxiv. 2018;
18. Holman L, Stuart-Fox D, Hauser CE. The gender gap in science: How long until women are equally represented? PLoS Biol. 2018;16(4):1–20.
19. King L, Mackenzie L, Tadaki M, Cannon S, Mcfarlane K, Reid D, et al. Diversity in geoscience : Participation , behaviour , and the division of scientific labour at a Canadian geoscience conference. 2018;415–40.
20. Ford HL, Brick C, Blaufuss K, Dekens PS. Gender inequity in speaking opportunities at the American Geophysical Union Fall Meeting. Nat Commun [Internet]. 2016;(2018):3–8. Available from: http://dx.doi.org/10.1038/s41467-018-03809-5
21. Lerback J, Hanson B. Journals invite too few women to referee. Nature. 2017;541.
22. Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. Science faculty ' s

294 subtle gender biases favor male students. PNAS. 2012;109(41):16474–9.
295 23. Porter AM, Ivie R. Women in Physics and Astronomy, 2019. AIP Stat Res Cent [Internet].
296 2019;45. Available from: www.aip.org/statistics
297 24. Unkovic C, Sen M, Quinn KM. Does encouragement matter in improving gender
298 imbalances in technical fields? Evidence from a randomized controlled trial. PLoS One.
299 2016;11(4):1–15.
300 25. Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. Science faculty's
301 subtle gender biases favor male students. Proc Natl Acad Sci. 2012;109(41):16474–9.
302 26. Breuning M, Gross BI, Feinberg A, Martinez M, Sharma R, Ishiyama J. Clearing the
303 Pipeline? Gender and the Review Process at the American Political Science Review. PS -
304 Polit Sci Polit. 2018;51(3):629–34.
305 27. Budden AE, Tregenza T, Aarssen LW, Koricheva J, Leimu R, Lortie CJ. Double-blind review
306 favours increased representation of female authors. Trends Ecol Evol. 2008;23(1):4–6.
307 28. Borusk R, Aarssen L, Budden A, Koricheva J, Leimu R, Tregenza TOM, et al. To Name or
308 Not to Name : The Effect of Changing Author Gender on Peer Review. Bioscience.
309 2009;59(11):985–9.
310 29. Bench SW, Lench HC, Liew J, Miner K, Flores SA. Gender Gaps in Overestimation of Math
311 Performance. Sex Roles. 2015;72(11–12):536–46.
312 30. Sarsons H, Xu G. Confidence men? Gender and confidence: Evidence among top
313 economists. Unpubl Manuscr [Internet]. 2015;1–26. Available from:
314 http://www.guoxu.org/docs/confidence.pdf
315 31. Morgan AC, Way SF, Clauset A. Automatically assembling a full census of an academic

field. PLoS One. 2018;13(8):1–18.

32. Pezzoni M, Mairesse J, Stephan P, Lane J. Gender and the Publication Output of Graduate Students : A Case Study. 2016;1–12.

33. Larivière V, Desrochers N, Mongeon P, Paul-hus A. Contributorship and division of labor in knowledge production. Soc Stud Sci. 2016;

34. Selenium [Internet]. Available from: https://www.selenium.dev/

35. Beautiful Soup.

36. genderize.io.

37. getGenders [Internet]. Available from: https://github.com/block8437/gender.py

Journal	Uncategorized	Male	Female	Unmatched initialed	Total
JGR Space Physics	644 (14.5%)	1114 (25.1%)	566 (12.7%)	2118 (47.7%)	4442
JGR Planets	58 (6.3%)	372 (40.6%)	166 (18.1%)	321 (35.0%)	917
JGR Earth Surface	45 (4.9%)	453 (49.2%)	171 (18.6%)	252 (27.4%)	921
Geophysical Research Letters	999 (12.5%)	3268 (40.9%)	1543 (19.3%)	2171 (27.2%)	7981
JGR Solid Earth	419 (14.3%)	1224 (41.8%)	618 (21.1%)	667 (22.8%)	2928
Nature Geoscience	77 (8.0%)	507 (52.5%)	217 (22.5%)	164 (17.0%)	965
Geology	153 (7.8%)	1145 (58.3%)	438 (22.3%)	229 (11.7%)	1965
GSA Bulletin	55 (7.5%)	435 (59.5%)	163 (22.3%)	78 (10.7%)	731
JGR Atmosphere	971 (19.6%)	1867 (37.7%)	1083 (21.9%)	1026 (20.7%)	4947
JGR Oceans	498 (16.1%)	1236 (39.9%)	723 (23.3%)	642 (20.7%)	3099

Geochimica & Cosmochimica Acta	392 (12.5%)	1443 (45.9%)	891 (28.3%)	418 (13.3%)	3144
Quaternary Science Reviews	181 (8.5%)	988 (46.4%)	646 (30.3%)	316 (14.8%)	2131
JGR Biogeosciences	154 (15.2%)	390 (38.5%)	310 (30.6%)	158 (15.6%)	1012
Total	4646 (13.2%)	14442 (41.0%)	7535 (21.4%)	8560 (24.3%)	35183

Table 1| Count for all names in journals scraped from January 2013 to April 2019. Proportions for uncategorized, male, female, and unmatched initialed names are shown in parentheses. Journal of Geophysical Research is abbreviated to JGR.