The Gender Gap in Scholarly Self-Promotion on Social Media

Hao Peng
University of Michigan https://orcid.org/0000-0002-5668-1680
Misha Teplitskiy
University of Michigan
Daniel Romero
University of Michigan https://orcid.org/0000-0002-8351-3463
Emőke-Ágnes Horvát (a-horvat@northwestern.edu)
Northwestern University

Article

Keywords:

Posted Date: June 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1765948/v1

The Gender Gap in Scholarly Self-Promotion on Social Media

Hao Peng[1,7], Misha Teplitskiy[1,2], Daniel M. Romero[1,3,4], Emőke-Ágnes Horvát[5,6,7]

[1]*School of Information, University of Michigan, Ann Arbor, MI 48109, USA*

[2]*Laboratory for Innovation Science at Harvard, Harvard University, Boston, MA, USA*

[3]*Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA*

[4]*Department of Electrical Engineering and Computer Science, University of Michigan, MI, USA*

[5]*School of Communication, Northwestern University, Evanston, IL 60208, USA*

[6]*McCormick School of Engineering, Northwestern University, Evanston, IL 60208, USA*

[7]*Northwestern University Institute on Complex Systems, Evanston, IL 60208, USA*

Self-promotion in science is ubiquitous but not exercised to the same extent by everyone. It is unclear whether there are gender differences in the frequency of self-promotion or the benefits individuals get from it. Here, we examine gender differences in scholarly self-promotion using 7M Tweet mentions of 539K research papers published in 2018 by 1.3M authors. Our analysis shows that female authors are significantly less likely than male authors to promote their papers, even after controlling for a number of important factors including journal impact, affiliation prestige, author productivity and number of citations, authorship position, number of coauthors, and research topics. The magnitude of the gender gap is more strongly associated with papers' journal impact factor than authors' affiliation prestige, previous productivity, or academic discipline. In particular, male scholars are 60% more likely than

comparable female scholars to self-promote papers published in journals with very high impact factor, whereas the difference is only 28% for papers in low impact journals. Although women self-promote less often overall, when they do, their papers receive slightly more mentions on Twitter. Our findings offer the first large-scale evidence for a gender gap in scholarly self-promotion online and show the circumstances under which the gap is most substantial, helping inform policy aimed at mitigating discrepancies in visibility and recognition.

Introduction

Traditional and social media have long played an important role in the dissemination of scientific findings[1–3]. An emerging line of literature suggests that the online visibility of scholarly papers amplifies their impact beyond academic audiences[4–6] and in the academy through citations[7–10], contributing increasingly to novel measures for scholarly evaluation[11,12]. Social media platforms, such as Twitter, are thus commonly used by scholars across disciplines to discuss ideas and disseminate findings[13–16]. Recent research shows consistent gender inequalities in the online dissemination of scholarly work[17], complementing mounting evidence for disparities in conventional scientific outcomes like citations, funding, and awards[18–20]. To better understand gender disparities in scholars' online visibility and inform the design of interventions[21–25], we need to focus on self-promotion, a behavior that directly speaks about one's strengths and achievements in professional contexts[26]. Promoting one's own research papers is a common and intuitive way to increase public attention to one's scholarly accomplishments. Here, we examine whether authors of different genders self-promote their research at different rates on social media and whether they get different returns on these efforts?

Publicizing one's own achievements is necessary for professional success in numerous settings including job interviews, salary negotiations, and career promotions[27,28]. Individuals who engage in self-promotion are doing so to be perceived as competent, ambitious, and confident which are traits associated with leadership skills[29,30]. Yet, women face a dilemma in self-promotion that may cause them to advocate for themselves less[31–33]. First, women risk a "backlash effect"[28,34,35],

in which audiences may penalize them for behavior that is incongruous with gender stereotypes or social expectations, such as women are modest and less assertive than men[36,37]. Indeed, empirical research has shown that women who self-promote can be seen as more arrogant and less likable than self-promoting men in many settings[38,39], and are judged to be gender nonconforming[30,31]. In addition to concerns about backlash, women may also self-promote less due to self-stereotyping in male-dominated domains, where they are under-represented and lack role models[40–43].

It is thus of little surprise that women are less likely than men to self-promote on online hiring platforms[44] and that female entrepreneurs' pitches are favored less by investors[45,46]. In the scientific domain, men have been found to be more likely than women to cite their papers[47–49] and present their research as novel and essential[50]. These findings suggest that female scientists may also be more conservative in the statement of their accomplishments than males in online self-promotion. Gender differences in academic self-promotion may be particularly consequential in the early stages of research dissemination, where small initial differences can accumulate into substantial disparities in coverage over time[51,52]. Further exacerbated by gender differences in media representations that favor men[53,54], inequalities in self-promotion might thus have a long-term effect on scholars' visibility and recognition, feeding into the persistent gender gap in science.

In this paper, we compiled a multi-disciplinary dataset of 539,345 research papers published in 2018 by 1,335,603 unique authors to examine the rate at which female vs. male scientists promote their research papers on social media. We focused on Twitter, which is the most commonly used platform for online science dissemination, accounting for 92% of all paper mentions on ma-

jor social media platform as tracked by Altmetric.com[55], the most comprehensive service to date for monitoring online posts about research papers. As each author of a paper has the option to promote it, we treated every (paper, author) pair as one of 2,375,419 observations of potential self-promotion cases. To link authors of papers to their Twitter accounts, if available, we designed and validated a heuristic to match author names to Twitter usernames and identify instances when an author mentioned their own paper in a tweet (i.e., they self-promoted). The frequency of self-promotion per paper is shown in SI Text A and SI Fig. S1.

We identified female and male authors based on their first names[56], categorizing 59% of the authors as male and 41% as female (unisex names and ambiguous cases were excluded from the analysis). While name-based binary gender may differ from how authors self-identify, author-declared gender is not available at the scale of our study. However, the gender inference employed here achieved a high level of agreement with male/female labels generated through a manual verification process and also with author self-reported gender labels from a recent journal submission dataset comprising 432,888 physicists. Finally, we collected metadata for all papers and their authors from the Microsoft Academic Graph[57], which enabled us to control for important factors that can affect self-promotion, such as the impact factor of the journal where the paper was published, the ranking of the author's institution, the author's previous productivity quantified by the number of their published papers and citation counts, the research fields associated with the paper, the byline position of the author, and the total number of co-authors. See Data and Methods for details about gender inference and its validation as well as the collected metadata.

The Gender Gap in Self-Promotion

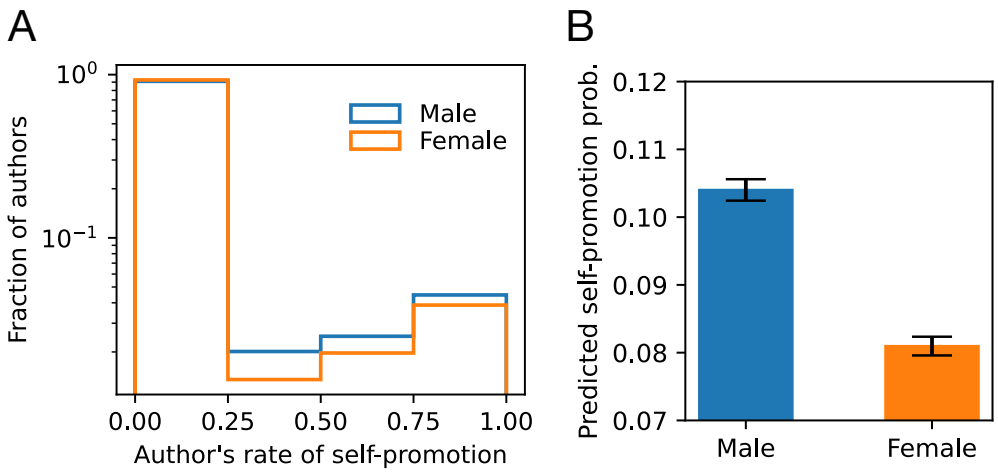

Figure 1: **Self-promotion by author gender.** **A**, Normalized histogram of the self-promotion rate per author, which is calculated as the fraction of an author's papers they have self-promoted. **B**, Predicted probability of self-promotion by gender after controlling for important confounding factors. Results show that women self-promote less on average than men even after accounting for journal impact, affiliation prestige, author productivity and number of citations, authorship position, number of coauthors, and research topics.

To examine whether there is a gender difference in how often authors promote their papers, we first calculated the overall self-promotion rate of each author, which is the fraction of papers they have ever promoted. People typically self-promote just once per paper. Fig. 1A shows that men tend to self-promote more often and there are more men than women who self-promote extensively. This finding also holds when focusing on authors with at least five publications in the data (SI Fig. S2), which suggests that the high self-promotion rate is not driven by authors with a few publications. Those with only one publication would have a rate of 100% if they self-promoted.

However, what fraction of their papers authors promote on average does not take into account differences in individual papers. To examine more nuanced factors related to self-promotion tendencies, we explored whether an author promotes differently each of their papers, and conversely whether a paper is self-promoted distinctly by its authors. Considering all possible (paper, author) instances that could have led to self-promotion had the author been a Twitter user at the time of publication, we found that men account for a substantially larger fraction than women (63.8% vs. 36.2%), which reflects a broader gender imbalance in publications and authorship[18,58]. Importantly, male scholars have an average self-promotion rate that is 21.9% higher than that of females (7.8% vs. 6.4%). This gender gap is universal across different byline positions and academic disciplines (SI Text B and SI Fig. S3).

Self-promotion is also likely to be influenced by factors related to impact and prestige. For instance, authors may tend to promote more of their papers published in high-impact journals, as people are more likely to share their higher achievements[50]. We find confirmatory evidence for this tendency in our data (SI Fig. S4A). Research has also revealed a gender imbalance across many prestige-related factors. For instance, women published fewer papers in leading journals[18,59], were placed at lower-ranked institutions[60], and were less likely to achieve tenure status[61].

We thus examined the association between author gender and self-promotion using a mixed effects logistic regression model (Data and Methods) that controlled for the random effects of each paper and several factors identified previously as related to visibility[17], including paper's journal impact, affiliation prestige, author's productivity and citations, number of coauthors, and

authorship position. We also controlled for research topics, as research fields vary substantially in gender representation in publishing and scholars' presence on Twitter[62–64], which can create an additional variability in self-promotion with regard to gender.

Based on the fitted model (see the regression coefficients in SI Table S1), we estimated the adjusted probability of self-promotion for both genders by setting all other variables at their median values across all observations. Fig. 1B shows that, a typical female author has an 8.1% chance of self-promoting her research, while the same quantity for a comparable male author is 10.4%, which is 28% higher than for females. This disparity also exists among female vs. male authors of the same paper, as our model accounts for the random effect of each paper.

Heterogeneity in Gender Disparities in Self-Promotion

To examine disparities across research fields, we fitted a separate mixed effects logistic regression model for each of the four broad disciplines of social, life, health, and physical sciences. Note that we still controlled for the fine-grained subject areas when fitting a model for each discipline. Fig. 2A shows the predicted probability of self-promotion for men and women in each broad discipline. The female coefficient in each regression model is significantly negative (SI Table S2), and the predicted mean self-promotion probability is higher for men than women across the board (Fig. 2A). However, there is variation in the typical self-promotion rate across disciplines. Although social scientists are on average about three times as likely to promote their research as scientists in the other three fields (0.21 vs. 0.07), the absolute gender disparity is universally similar

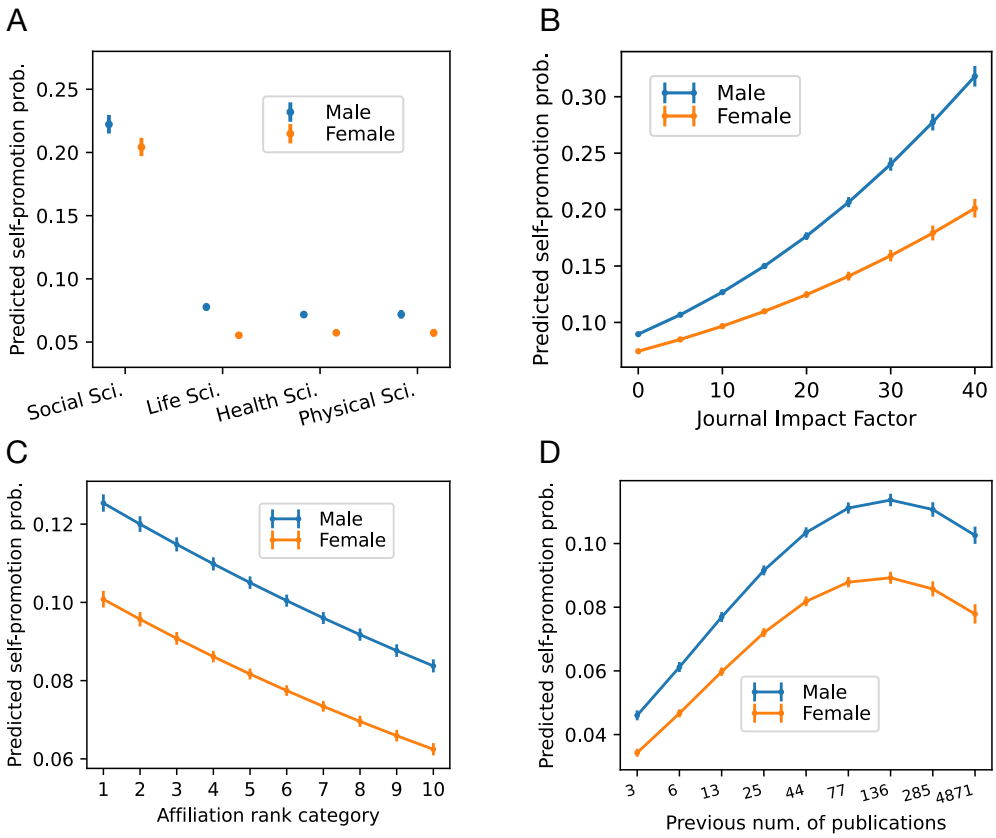

Figure 2: Predicted probability of self-promotion by gender. **A**, Variation across four broad disciplines. While the gender gap in predicted self-promotion probability is of similar magnitude across four fields, social scientists have a high average self-promotion probability regardless of gender, which leads to the smallest relative gender gap between the two groups. **B-D**, Self-promotion as a function of journal impact, affiliation prestige, and author productivity. Journal impact factor (**B**) is provided by the Web of Science. We categorized (deciled) in the regression author's affiliation rank (**C**, a smaller bin indicates a higher rank category) and authors' previous number of publications (**D**, a smaller bin indicates a less productive category). Error bars indicate 95% bootstrapped confidence intervals.

across disciplines, suggesting that men are self-promoting more often than women in science in general. However, the higher overall rate of self-promotion also makes the relative gender gap in the social sciences the smallest (10% vs. 40% in the other three disciplines), indicating that both male and female scholars in this field are more active in promoting their research compared to other disciplines.

To explore how different factors affect the magnitude of gender disparities in self-promotion, we included in the model a separate interaction term between gender and journal impact, affiliation prestige, and author productivity (see raw gender differences in SI Fig. S4), and calculated the predicted probability of self-promotion for both genders as we did in Fig. 1B.

Publishing in top-tier journals can boost authors' confidence and therefore may reduce the gender disparity in self-promotion, leading to authors of both genders becoming more active in sharing the research. In contrast, past research shows that high-achieving women can elicit more pushback for their success as they violate gender expectations [65]. This may cause female scholars to self-promote their high-impact papers less. These competing hypotheses raise the question whether journal impact decreases or increases the gender disparity in self-promotion rate.

Fig. 2B shows that authors of both genders are more likely to share research published in higher impact than less prestigious venues. However, this effect is less pronounced for females, resulting in an even larger gender disparity in case of papers published in top-tier venues. Considering papers published in journals with impact factor 40, male scholars are 60% more likely to self-promote than female scholars (0.32 vs. 0.20), whereas for low-tier publications (impact factor

below 5), the same probability is only 28% higher for men than for women (0.09 vs. 0.07).

Fig. 2C shows that authors affiliated with prestigious institutions self-promote much more often than those from lower-ranked affiliations. Specifically, authors from the highest prestige decile are about 50% more likely to self-promote than those from the lowest-tier affiliations. However, this effect applies nearly equally to both genders, resulting in a similar size of gender gap in self-promotion probabilities across the institution hierarchy: regardless of affiliation ranks, men's self-promotion probability is, on average, 2.5 percentage points higher than women's.

As shown in Fig. 2D, there is a quadratic relationship between author's research experience and the likelihood of self-promotion. Throughout a scholar's career, their self-promotion probability increases up to the mid-career stage, after which the probability starts to decrease. Mid-career scholars are much more likely to self-promote than early- and late-career scholars. This pattern is similar for both genders. One possible reason is that junior researchers spend most of the time conducting research to build their portfolio, while the task of marketing papers is often taken on by more senior scholars[66,67]. However, as senior authors further make progress in their careers and become more established, they have less time and incentives to self-promote. The largest gender disparity occurs among mid-career authors, where both male and female scholars are at the peak of their self-promotion probabilities (0.114 vs. 0.088).

These results suggest that there are well-defined and contrasting gender patterns in self-promotion on Twitter, with men and women adopting slightly different tactics in sharing their own research. Specifically, men are much more active in self-promotion than women when they are in

the mid-career stage and publish in high-impact journals. The findings illustrate the subtle nature of gender differences in scholarly self-advertising on social media and provide new insights into the gender inequities in scientists' online visibility and success.

Are There Disparities in the Return on Self-Promotion?

Self-promotional tweets are expected to attract attention to one's paper. However, we know little about the effect of self-promotion on a paper's overall popularity and whether there is a gender difference in the return associated with self-promotion. We set out to measure these effects.

To model the right-skewed distribution of a paper's popularity quantified by the total number of tweets, we fitted a negative binomial regression model[68] to this dependent variable. In our data, female scholars' papers received less attention than male scholars' papers (average total number of tweets: 16.8 vs. 18.2; $p < 0.001$), which is consistent with previous findings[17]. To examine how self-promotion impacts the total number of tweet mentions of a paper differently for men and women, we included in the model an interaction term between author gender and self-promotion. The control variables in this analysis are defined similarly as in modeling the probability of self-promotion (Data and Methods).

As shown in Table 1, self-promotion can indeed significantly increase a paper's total number of tweet mentions. Surprisingly, women are slightly advantaged in receiving online attention for their papers (the "Female" coefficient is positive and significant), after controlling for important confounding factors that are associated with attention. Moreover, although women self-promote

Table 1: Coefficients of all variables in a negative binomial regression model that predicts the total number of tweets for each (author, paper) observation. Subject area controls are omitted here for space consideration. See full regression table in SI Table S5. Significance levels: *** $p<0.001$, ** $p<0.01$, and * $p<0.05$.

Female	0.015	***
Self promotion = True	1.569	***
Female x (Self promotion = True)	0.038	***
First position	0.009	*
Middle position	0.270	***
Solo author	−0.196	***
Author pub. count category	−0.077	***
Affiliation rank category	−0.024	***
Affiliation location = International	−0.102	***
Number of authors	0.0001	***
Journal impact	0.093	***
Author log citations	0.055	***
Constant	1.675	***

less often than men, when they do, they get on average more overall attention to their papers (the small but positive coefficient for the interaction term). The model's adjusted prediction of a paper's total number of tweets reveals that, without self-promotion, females' higher popularity over males is very limited on average (the predicted total number of tweets: 7.84 vs. 7.73). However, with self-promotion, women have a 5.4% higher popularity than men (the predicted total number of tweets: 39.13 vs. 37.13).

To test the robustness of this finding, we ran three additional tests. (1) We checked that the result is consistent when fitting the same model to the subject sample from our data that are involved in self-promotion. For this test, we considered only (paper, author) pairs that contain self-promotion, dropped the interaction term between author gender and self-promotion, and additionally controlled for the author's follower count on Twitter (SI Table S6). (2) As our model treats each (paper, author) pair as the unit of analysis, a paper's total number of tweets might be affected by the self-promotion benefit from all coauthors. To better measure the gender effect in the presence of teams, we fitted a separate model to solo-authored papers (dropping controls for the number of authors and authorship position) and obtained similar results (SI Table S7). (3) We tested that the finding is also robust when defining self-promotion as posting about the paper at least once within one day after paper publication. This analysis suggests that the gender difference in return even applies to timely self-promotion behaviors (SI Table S8).

In summary, our analyses point to the surprising finding that women have a slight advantage over men in the return on self-promotion of research papers on social media.However, establishing

a causal relationship between self-promotion and online visibility requires further experimental investigations. This finding suggests that a key component to closing the gender gap in visibility on social media is enabling that female scholars can self-promote more.

Discussion

Our analysis based on 1.3M authors and 7M tweet mentions of their published research papers shows that scholarly self-promotion on social media is highly dependent on gender. Female scholars are significantly less likely to advertise their papers on Twitter than their male colleagues. This association persists even after controlling for a number of important factors, including journal impact, affiliation prestige and location, author productivity and citations, authorship position, and research areas. The disparity also occurs in different types of self-promotion, including *original tweets* and *retweets* (SI Fig. S5; see details about distinguishing between the two types of self-promotion in Data and Methods).

Our model does not control for authors' presence on Twitter. Thus, a lack of self-promotion could be due to an author who has a Twitter account but chooses not to (re)tweet about their paper, or an author who is not on Twitter at all. While these are two different situations, they are both cases where the author does not engage in and benefit from self-promotion on the platform. The observed gender gap in self-promotion is unlikely to be explained by women's under-representation on Twitter, as past research shows that female academics are at least as likely as men to be on this social media site[62]. As a robustness check, we obtained consistent results (SI Table S3) when

restricting the analysis to the subset of authors who have self-promoted at least once in our data (therefore have a Twitter account). Note that this subset of the data also contains (author, paper) pairs that did not self-promote since not all authors self-promote all their papers.

Regardless of gender, we find that self-promotion behavior varies substantially across disciplines, affiliations, journal impact, and author productivity. Social scientists self-promote much more often than scholars from life, physical, and health sciences. In general, scholars are much more likely to self-promote when they publish papers in top-tier journals, are affiliated with prestigious institutions, and are at their mid-career stages.

The gender disparity is the largest for papers published in top-tier journals, suggesting that while publishing in high-impact journals increases the willingness to self-promote for both genders, this effect is stronger for male authors. The same effect can be interpreted as men are more sensitive to associating their names with papers published in lower-tier journals and women are held back in promoting their top-tier papers. However, the gender gap is more or less stable across disciplines, affiliation ranks, and author productivity.

We further find that, while all authors receive higher social media attention when they self-promote their papers, this increase in popularity is somewhat higher for female than male scholars. These results hold even when we consider mentions of the paper by scientists only or non-scientists only (SI Tables S9-S10). This differential return on self-promotion might be due to different language styles used by men and women[69]. A preliminary sentiment analysis of the textual content involved in self-promotion shows that women use more positive language when sharing their re-

search than men, whereas men are more neutral in their language than women (SI Text C and SI Fig. S6). If positive content attracts more engagement on Twitter, this may lead to women receiving more online attention for their self-promoted papers. Regardless of the underlying mechanisms, the finding that female scholars are at least as effective as male scholars in attracting attention on Twitter to their self-promoted papers suggests that enabling female scholars to be more active in self-promotion (or male scholars less) may be one solution for reducing gender disparities in scholars' online success[17]. Nevertheless, it is unclear whether and how a systematic shift in women's self-promotion frequency would change subsequent attention to their papers.

The gender difference in self-promotion is consistent with gender gaps observed offline, such as gaps in self-citation[47] and research presentation[50]. This universality in female scientists' lower tendency to self-promote points to persistent factors preventing them from doing so. It is critical that future research unpacks these because small differences in the initial stage of research dissemination may grow into large disparities in scholars' ultimate visibility across different online platforms and recognition via citations and awards.

Although our observational study does not uncover the reasons behind the observed gender differences in the rate of self-promotion, several of our analyses point to a process that likely involves multiple mechanisms proposed in the literature, laying the foundations for further work in this space. An estimated 85% of women and girls globally have experienced some form of online harassment and abuse[70]. Harassment is most common on social media sites with 61% of women considering it a major problem[71]. Female scholars and graduate students, especially those

who are also racial minorities, are prone to be exposed to negative events including defamatory statements, stalking, and misogynistic and gender-oriented attacks[72–75]. Such attacks and abuse can substantially discourage self-promotion by female scholars, particularly up-and-coming researchers. In online science dissemination, women may be more conservative in self-promotion due to fear of backlash[28,35] and self-stereotyping[40]. In support of this, we find that the gender gap in self-promotion is smaller in the form of retweeting others' tweets that promote one's work than in the form of tweeting on their own (SI Fig. S5), likely because the former case reduces the exposure associated with self-promotion. Furthermore, women are not self-promoting their high-impact research as much as men do, possibly out of fear of getting more backlash for advertising their higher achievements[34,65]. Other mechanisms may also play an effect, such as that women might be more likely to adopt a modest self-presentation style than men[36]. In support of this, we analyzed the language used in self-promotion. We found that men more often use "novel", "excellent", "remarkable", and "unprecedented" than women, whereas women more often use "amazing", "creative", "supportive", and "inspiring" when promoting their research (SI Text C and SI Fig. S7). Examining these mechanisms is a fruitful avenue for future research.

Our study is not without limitations. First, our analysis focused on papers published in 2018, which might not generalize to other time periods if gender representation and usage of the platform changed over time. Second, we limited our study to a single social media platform, Twitter. Although it is the largest platform based on mentions of scientific papers, it would be interesting to extend the investigation to other social media sites. Third, we used name-based inferred gender as a proxy of authors' gender identity. Although our validation shows that there

is a high degree of agreement between the predicted attribute and manually verified or even self-reported gender, the inferred identity and the true identity may differ in some cases. This issue might be more severe for authors with East Asian names because many of them are gender neutral. Our main results about self-promotion are similar for authors with non-East Asian names (SI Table S4). Moreover, we are lacking accurate ways to identify non-binary genders[76,77] and make claims about their self-promotion rates. Fourth, our study requires matching authors to Twitter users. We addressed this problem using a name-based matching algorithm. Although our validation shows that our matching process achieves a very high accuracy, the algorithm is not without any errors. Future work can develop more sophisticated methods to address this issue.

Our study offers novel insights into scholars' self-promotion behaviors on social media in general and based on scientists' gender, enriching our understanding of the broader gender inequity in scientists' online visibility. Our findings have policy implications for online science dissemination. For instance, to restore gender parity in scholars' online visibility, institutions could allocate resources, provide training, dedicated time, and support in case of online harassment to facilitate more self-promotion by female scholars. This effort should be especially rewarding since women's self-promotion is associated with a significant increase in the online visibility of their work, providing younger generations with successful role models.

Data and Methods

Altmetric database Our data are based on the most complete record of research papers' online mentions, maintained by Altmetric.com[55]. This service has been tracking the online mentions of research outputs since 2011 in different platforms including news media and social media such as Twitter and Facebook. We accessed the database (referred to as "Altmetric" hereafter) on Oct 8, 2019. Altmetric matches the online attention to papers based on their unique identifiers, such as the Digital Object Identifier (DOI), PubMed ID, and arXiv ID. Utilizing these identifiers, it also collapses the attention for different versions of each paper into a single unique record[78]. This ensures that the data contains the complete mentions of each paper.

We focused on all 1,218,710 research papers (before dropping observations with missing values for all control variables) published in 2018 in the database. We obtained all their mentions in the largest platform, Twitter, which consists of about 74% of all mentions (posts) and 92% of all social media mentions in the entire database[79].

Due to Altmetric's data license agreement with Twitter, the dataset contains only the tweet ID for each tweet. We thus collected all tweets using the Twitter API. Due to account deletion or changes in the privacy settings by some Twitter users, we successfully retrieved about 90% of all tweets. The Altmetric data also provides metadata for each paper, such as the DOI, publication year, publication venue, research topics. The disciplinary catalog uses 26 Scopus Subject Areas, which belong to 4 broad disciplines including Social Sciences, Life Sciences, Physical Sciences, and Health Sciences[80]. The classification was performed by in-house experts based on the aim and

scope of the content a journal publishes[80].

Microsoft Academic Graph We used the Microsoft Academic Graph (MAG) database[57,81] (accessed on June 01, 2019) to retrieve other metadata for each paper based on the DOI. We obtained essential author information for each paper including author name, author affiliation, affiliation location, affiliation rank. We also counted author's previous number of citations and publications up to the publication year of each paper, which reflects author's previous productivity when deciding whether to self-promote or not, assuming authors tend to self-promote soon after the publication of a paper. MAG leverages data mining and artificial intelligence techniques to address author conflation and disambiguation, which ensures that the author's number of publications is counted accurately[57,82]. A paper can have multiple authors and we track whether each of them posted about the paper on Twitter. We, therefore, used each (paper, author) pair as an observation in the analysis. In our final dataset, after dropping missing values for all control variables, there were 2,380,098 observations for 539,848 papers, which were mentioned in about 7 million tweets.

Gender prediction We used Ford's algorithm[56] to infer the gender of a author based on their first name. The algorithm returns, for a given name, one of 4 categories: Female, Male, Unisex, Unknown. A name is predicted as "Female" ("male") if it is used for women (men) at least twice as frequently as for men (women), based on data from national statistics institutes[83]. Otherwise, the name is labeled as "Unisex" (gender-neutral). If the name is not found in the database, it is labeled as "Unknown". The percentage of observations for Male, Female, Unknown, and Unisex is 49%, 28%, 17%, 6%, respectively. In the final dataset, we excluded "Unknown" and "Unisex"

categories (we obtained consistent results when including the two categories). We used "Male" as the reference category in the regressions.

Validation with manually verified gender

We evaluated the performance of this algorithm based on a random sample of 100 authors in our data, for which we manually labelled their gender. In the labeling process, the author gender was determined based on their pronouns and profile pictures displayed on their personal websites, institutional directories, and Wikipedia pages found via a Google search of author names. We used the author's affiliation to disambiguate multiple authors with the same name. If the gender could not be verified, it was labeled as "Unknown". There were 19 females, 57 males, and 24 authors with unknown gender in the labelled sample. Based on the set of 76 authors with confirmed gender, the algorithm achieved an accuracy of $F1_m = 0.91$ and $F1_f = 0.88$. A different algorithm, the Genderize API[50,84–86], produced a similar result.

Validation with author self-reported gender

We also validated the accuracy of the gender prediction using author self-identified gender labels in the data provided by IOP Publishing (`https://ioppublishing.org/`), which is an academic publishing company specialized in the field of physics.

Each author has self-reported their gender and country of residence such as China, India,

U.S., Canada, Australia, etc. In our evaluation, we focused on authors with self-reported gender as either male or female, and whose names were predicted as either male or female, as our actual analysis was focused on authors with distinctly predicted gender labels. There are 432,888 authors submitting to 62 journals in this data. Here we list authors from China as a separate group because Chinese names typically do not encode clear gender signal when written in English characters[87]. For instance, we found that the vast majority of names (475 such names in total) with both male and female as self-reported gender are from China. The prediction accuracy for Chinese names is lower than non-Chinese names, and the overall F1 score is close to 0.9 (SI Table S11). However, this is likely an underestimation because China accounts for a much larger fraction of the papers in the IOP data than in the Altmetric data.

Robustness check on non-East Asian names

Since the gender prediction is typically less accurate for East Asian names[88], we repeated our analysis by excluding observations whose author names were predicted to be East Asian ethnicities using the Ethnea API [89]. For a specific name, *Ethnea* assigns the ethnicity probabilities based on matched authors in the PubMed database. In the case of two or more predicted ethnicities, we took the one with the highest probability. *Ethnea* predicts 26 individual ethnicities. We categorized 7 of them as East Asian, including *Chinese*, *Indonesian*, *Japanese*, *Korean*, *Mongolian*, *Thai*, *Vietnamese*. We obtained consistent results on the subset of the data (2M observations) with non-East Asian names (SI Table S4).

Detecting self-promotion among a paper's tweet mentions Each paper has a list of tweets that mention the paper, with each tweet containing the twitter handle and the screen name of the user (referred to as "tweet names" hereafter). We defined self-promotion as the author posting a tweet sharing the unique identifier of their paper[78], such as the DOI, PubMedID, arXiv ID, etc. Self-promotion on Twitter comes in different forms, which can manifest in the type of the tweet that shares the research, including (i) an original tweet, or (ii) a retweet. The two types of self-promotion differ in how direct the promotion is: the original-tweet-based promotion comes from the authors themselves whereas the retweet-based promotion originates from others (e.g., an author A retweets a tweet from others sharing A's paper). Self-promotion in the form of posting an original tweet potentially indicates a stronger intention to advertise one's paper than that based on a retweet, which is more indirect in the nature of promotion.

We constructed a binary dependent variable indicating whether the author self-promoted the paper or not, based on both retweets and original tweets. We determined if an author was among the users who tweeted the paper using string matching between names.

There is no perfect method to match author names with tweet names because scholars can use any string as their handle or screen name on Twitter. We thus adopted a simple "containment-matching" approach that searched the author name in tweet names—if either the first name or the last name string was contained in the user handle or screen name (lowercased), we considered the user to be the author of the paper. In case of multiple matches, we used the one with the highest fuzzy matching score[90].

We validated this method using a random sample of 100 papers (each with at least one tweet mentioning it) with manual verification. Due to having multiple authors per paper, there were 521 (paper, author) pairs in the manual labelling process. For each observation, we verified the author against all tweets of the paper to check if the author was among the tweet users. This method achieved an initial F1 score of 0.85 (precision: 0.77 and recall: 0.95).

We refined this method through an iterative process by experimenting with various new heuristics to address false predictions and evaluating its performance based on an independent manual verification process at each iteration. In the final version, we used "containment-matching" only if the tweet names are single-token string and the author's first name (or last name) had at least 4 characters; otherwise, we used "token-matching', i.e., the first name or the last name should be matched to the tokens of tweet names (split by space or underscore). This final heuristic achieved an out-of-sample precision of 0.95 and a recall of 1.00 (F1 score: 0.97).

Modelling the likelihood of self-promotion We used a mixed effects logistic regression model[91] to estimate the probability that an author self-promoted a paper on Twitter as a function of their gender, while controlling for the following fixed and random effects variables:

- *Journal impact factor*: Papers published in high impact journals may be more likely to be shared by their authors. We obtained the journal impact factor from The Web of Science (2018 version).
- *Rank of authors' affiliation*: Authors from prestigious institutions may be more likely to

self-promote. We thus considered the rank of their affiliations provided in the MAG. When an author has multiple affiliations in a paper, we used the one with the highest rank. The rank value is log-transformed in the MAG. We therefore categorized the rank values into ten equally-sized bins (a smaller bin indicates a higher rank category).

- *Authors' number of publications prior to 2018*: Authors' career stage and previous research experience can influence their likelihood of self-promoting. To measure this factor, we counted each author's total number of publications before 2018, using all papers indexed in MAG. We also categorized this numerical variable into ten equally-sized bins due to the non-linearity of this variable, and to reduce noise and outliers.

- *Authors' number of citations prior to 2018*: Authors' citation counts can impact both their likelihood of self-promotion and the return associated with self-promotion. We counted each author's total number of citations before 2018 using the MAG data. We log-transformed this highly skewed variable to make it more interpretable in the model.

- *Number of authors*: We counted the number of authors in each paper. Having more coauthors in a paper could potentially impact an author's likelihood of self-promoting the paper on their own.

- *Byline position*: Different authors often play different roles in multi-author projects. For example, authors who play a supportive role in the project may self-promote less frequently. This variation is often captured by the authorship position in the paper. We thus controlled for the position of an author with four categories: (1) first position, (2) middle position, (3)

last position, (4) solo author. The last position was used as the reference category in the regression.

- *Affiliation location*: We inferred the country of the author's institution using the latitude and longitude information in the MAG. There were two categories: (1) U.S., (2) international. We used "U.S." affiliation as the baseline to control for the fact that Twitter is a U.S. social media platform that is more likely to be adopted by authors based in the U.S. Note that 4679 observations with unknown affiliation location information were excluded in the analysis. When an author had multiple affiliations with at least one located in the U.S., we classified them as "U.S." In the regression, we treated "U.S." as the reference category.

- *Research fields*: Not all scholars employ Twitter as a channel to share their research, and scientists' representation on Twitter varies across disciplines[63]. To control for field-specific effects, we used the 26 Scopus Subject Areas. Each subject area was treated as a fixed variable in the regression, whose value was coded as 1 if the paper was assigned that subject (0 otherwise). Note that a paper can belong to multiple subject areas.

- *Paper random effects*: Individual papers have different degrees of newsworthiness (e.g., biomedical papers have much more online coverage than papers from other disciplines[92]). Different papers may vary in the likelihood of being shared on social media by their authors. Gender representation online also varies across disciplines[17]. To capture such paper-level variations, we added random effects for each paper in the model.

Since there was a quadratic relationship between author productivity and self-promotion rate

in the data (SI Fig. S4), we included a second order polynomial term for author productivity in the model. We treated all other variables as linear terms in the regression.

1. Editorial. Social media for scientists. *Nature Cell Biology* **20**, 1329 (2018).

2. Peters, H. P. Gap between science and media revisited: Scientists as public communicators. *Proceedings of the National Academy of Sciences* **110**, 14102–14109 (2013).

3. Rowlands, I., Nicholas, D., Russell, B., Canty, N. & Watkinson, A. Social media use in the research workflow. *Learned Publishing* **24**, 183–195 (2011).

4. Eagleman, D. M. Why public dissemination of science matters: A manifesto. *Journal of Neuroscience* **33**, 12147–12149 (2013).

5. Cronin, B. & Sugimoto, C. R. *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (MIT Press, 2014).

6. Sugimoto, C. R., Work, S., Larivière, V. & Haustein, S. Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology* **68**, 2037–2062 (2017).

7. Costas, R., Zahedi, Z. & Wouters, P. Do "altmetrics" correlate with citations? extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology* **66**, 2003–2019 (2015).

8. Dehdarirad, T. Could early tweet counts predict later citation counts? a gender study in life sciences and biomedicine (2014–2016). *PLOS ONE* **15**, e0241723 (2020).

9. Klar, S., Krupnikov, Y., Ryan, J. B., Searles, K. & Shmargad, Y. Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work. *PLOS ONE* **15**, e0229446 (2020).

10. Luc, J. G. *et al.* Does tweeting improve citations? one-year results from the tssmn prospective randomized trial. *The Annals of Thoracic Surgery* **111**, 296–300 (2021).

11. Kwok, R. Research impact: Altmetrics make their mark. *Nature* **500**, 491–493 (2013).

12. Hicks, D., Wouters, P., Waltman, L., De Rijcke, S. & Rafols, I. Bibliometrics: the leiden manifesto for research metrics. *Nature News* **520**, 429 (2015).

13. Priem, J. & Costello, K. L. How and why scholars cite on twitter. *Proceedings of the American Society for Information Science and Technology* **47**, 1–4 (2010).

14. Darling, E., Shiffman, D., Coté, I. & Drew, J. The role of twitter in the life cycle of a scientific publication. *Ideas in Ecology and Evolution* **6** (2013).

15. Morello, L. We can# it: social media is shaking up how scientists talk about sexism and gender issues. *Nature* **527**, 148–152 (2015).

16. Gero, K. I., Liu, V., Huang, S., Lee, J. & Chilton, L. B. What makes tweetorials tick: How experts communicate complex topics on twitter. *Proceedings of the ACM on Human-Computer Interaction* **5**, 1–26 (2021).

17. Vasarhelyi, O., Zakhlebin, I., Milojevic, S. & Ágnes Horvát, E. Gender inequities in the online dissemination of scholars' work. *PNAS* (2021).

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

19. Way, S. F., Larremore, D. B. & Clauset, A. Gender, productivity, and prestige in computer science faculty hiring networks. In *Proceedings of the 25th International Conference on World Wide Web*, 1169–1179 (2016).

20. Oliveira, D. F., Ma, Y., Woodruff, T. K. & Uzzi, B. Comparison of national institutes of health grant amounts to first-time male and female principal investigators. *JAMA* **321**, 898–900 (2019).

21. Moss-Racusin, C. A. *et al.* Scientific diversity interventions. *Science* **343**, 615–616 (2014).

22. Matias, J. N., Szalavitz, S. & Zuckerman, E. Followbias: supporting behavior change toward gender equality by networked gatekeepers on social media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1082–1095 (2017).

23. Pietri, E. S. *et al.* Addressing unintended consequences of gender diversity interventions on women's sense of belonging in stem. *Sex Roles* **80**, 527–547 (2019).

24. Fortin, J., Bartlett, B., Kantar, M., Tseng, M. & Mehrabi, Z. Digital technology helps remove gender bias in academia. *Scientometrics* **126**, 4073–4081 (2021).

25. Langrock, I. & González-Bailón, S. The gender divide in wikipedia: Quantifying and assessing the impact of two feminist interventions. *Journal of Communication* (2022). URL `https://doi.org/10.1093/joc/jqac004`. Jqac004,

https://academic.oup.com/joc/advance-article-pdf/doi/10.1093/joc/jqac004/42540403/jqac004.pdf.

26. Schlenker, B. R. *Impression management* (Monterey, CA: Brooks/Cole, 1980).

27. Laschever, S. & Babcock, L. *Women Don't Ask: Negotiation and the Gender Divide* (Princeton University Press, 2003).

28. Moss-Racusin, C. A. & Rudman, L. A. Disruptions in women's self-promotion: The backlash avoidance model. *Psychology of Women Quarterly* **34**, 186–202 (2010).

29. Stevens, C. K. & Kristof, A. L. Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology* **80**, 587 (1995).

30. Rudman, L. A. Self-promotion as a risk factor for women: the costs and benefits of counter-stereotypical impression management. *Journal of Personality and Social Psychology* **74**, 629 (1998).

31. Janoff-Bulman, R. & Wade, M. B. The dilemma of self-advocacy for women: Another case of blaming the victim? *Journal of Social and Clinical Psychology* **15**, 143–152 (1996).

32. Wade, M. E. Women and salary negotiation: The costs of self-advocacy. *Psychology of Women Quarterly* **25**, 65–76 (2001).

33. Exley, C. L. & Kessler, J. B. The gender gap in self-promotion. *The Quarterly Journal of Economics* (2022). URL https://doi.org/10.1093/qje/qjac003.

34. Phelan, J. E. & Rudman, L. A. Prejudice toward female leaders: Backlash effects and women's impression management dilemma. *Social and Personality Psychology Compass* **4**, 807–820 (2010).

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

36. Daubman, K. A., Heatherington, L. & Ahn, A. Gender and the self-presentation of academic achievement. *Sex Roles* **27**, 187–204 (1992).

37. Heatherington, L. *et al.* Two investigations of "female modesty" in achievement situations. *Sex Roles* **29**, 739–754 (1993).

38. Hagen, R. L. & Kahn, A. Discrimination against competent women. *Journal of Applied Social Psychology* **5**, 362–376 (1975).

39. Rudman, L. A. & Glick, P. Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues* **57**, 743–762 (2001).

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

41. Josephs, R. A., Markus, H. R. & Tafarodi, R. W. Gender and self-esteem. *Journal of Personality and Social Psychology* **63**, 391 (1992).

42. Bleidorn, W. *et al.* Age and gender differences in self-esteem—a cross-cultural window. *Journal of Personality and Social Psychology* **111**, 396 (2016).

43. Wang, Y. & Horvát, E. Gender differences in the global music industry: Evidence from musicbrainz and the echo nest. In *Proceedings of the Thirteenth International Conference on Web and Social Media, ICWSM 2019, Munich, Germany, June 11-14, 2019.*, 517–526 (2019).

44. Altenburger, K., De, R., Frazier, K., Avteniev, N. & Hamilton, J. Are there gender differences in professional self-promotion? an empirical case study of linkedin profiles among recent mba graduates. In *Proceedings of the International AAAI Conference on Web and Social Media* (2017).

45. Brooks, A. W., Huang, L., Kearney, S. W. & Murray, F. E. Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences* **111**, 4427–4431 (2014).

46. Horvát, E. o.-A. & Papamarkou, T. Gender differences in equity crowdfunding. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 5 (2017).

47. King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J. & West, J. D. Men set their own cites high: Gender and self-citation across fields and over time. *Socius* **3**, 2378023117738903 (2017).

48. Andersen, J. P., Schneider, J. W., Jagsi, R. & Nielsen, M. W. Meta-research: Gender variations in citation distributions in medicine are very small and due to self-citation and journal prestige. *eLife* **8**, e45374 (2019).

49. Azoulay, P. & Lynn, F. B. Self-citation, cumulative advantage, and gender inequality in science. *Sociological Science* **7**, 152–186 (2020).

50. Lerchenmueller, M. J., Sorenson, O. & Jena, A. B. Gender differences in how scientists present the importance of their research: observational study. *British Medical Journal* **367** (2019).

51. Lieberman, M. B. & Montgomery, D. B. First-mover advantages. *Strategic Management Journal* **9**, 41–58 (1988).

52. Bol, T., de Vaan, M. & van de Rijt, A. The matthew effect in science funding. *Proceedings of the National Academy of Sciences* **115**, 4887–4890 (2018).

53. Atir, S. & Ferguson, M. J. How gender determines the way we speak about professionals. *Proceedings of the National Academy of Sciences* **115**, 7278–7283 (2018).

54. Peng, H., Teplitskiy, M. & Jurgens, D. Author mentions in science news reveal wide-spread ethnic bias. *arXiv:2009.01896* (2020).

55. Altmetric. `https://www.altmetric.com/` (2021).

56. Ford, D., Harkins, A. & Parnin, C. Someone like me: How does peer parity influence participation of women on stack overflow? In *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 239–243 (IEEE, 2017).

57. Wang, K. *et al.* A review of microsoft academic services for science of science studies. *Frontiers in Big Data* **2**, 45 (2019).

58. West, J. D., Jacquet, J., King, M. M., Correll, S. J. & Bergstrom, C. T. The role of gender in scholarly authorship. *PLOS ONE* **8**, e66212 (2013).

59. Editorial. Women feature only rarely as first or last authors in leading journals. *Nature* **555**, 691 (2018).

60. Smart, J. C. Gender equity in academic rank and salary. *The Review of Higher Education* **14**, 511–525 (1991).

61. Jena, A. B., Khullar, D., Ho, O., Olenski, A. R. & Blumenthal, D. M. Sex differences in academic rank in us medical schools in 2014. *JAMA* **314**, 1149–1158 (2015).

62. Bowman, T. D. Differences in personal and professional tweets of scholars. *Aslib Journal of Information Management* (2015).

63. Ke, Q., Ahn, Y.-Y. & Sugimoto, C. R. A systematic identification and analysis of scientists on twitter. *PLOS ONE* **12**, e0175368 (2017).

64. Costas, R., Mongeon, P., Ferreira, M. R., van Honk, J. & Franssen, T. Large-scale identification and characterization of scholars on twitter. *Quantitative Science Studies* **1**, 771–791 (2020).

65. Cooper, M. For women leaders, likability and success hardly go hand-in-hand. *Harvard Business Review* **30** (2013).

66. Wren, J. D. *et al.* The write position: A survey of perceived contributions to papers based on byline position and number of authors. *EMBO Reports* **8**, 988–991 (2007).

67. Sekara, V. *et al.* The chaperone effect in scientific publishing. *PNAS* **115**, 12603–12607 (2018).

68. Hilbe, J. M. *Negative binomial regression* (Cambridge University Press, 2011).

69. Bamman, D., Eisenstein, J. & Schnoebelen, T. Gender identity and lexical variation in social media. *Journal of Sociolinguistics* **18**, 135–160 (2014).

70. Crockett, C. & Vogelstein, R. Launching the global partnership for action on gender-based online harassment and abuse. *The White House* (2022). URL https://www.whitehouse.gov/gpc/briefing-room/2022/03/18/launching-the-global-partnership-for-action-on-gender-based-online-haras

71. Vogels, E. A. The state of online harassment. *Pew Research Center* **13** (2021).

72. Vitak, J., Chadha, K., Steiner, L. & Ashktorab, Z. Identifying women's experiences with and strategies for mitigating negative effects of online harassment. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1231–1245 (2017).

73. Veletsianos, G., Houlden, S., Hodson, J. & Gosse, C. Women scholars' experiences with online harassment and abuse: Self-protection, resistance, acceptance, and self-blame. *New Media & Society* **20**, 4689–4708 (2018).

74. Samer, C., Lacombe, K. & Calmy, A. Cyber harassment of female scientists will not be the new norm. *The Lancet Infectious Diseases* **21**, 457–458 (2021).

75. Nogrady, B. *et al.* 'i hope you die': how the covid pandemic unleashed attacks on scientists. *Nature* **598**, 250–253 (2021).

76. Tannenbaum, C., Ellis, R. P., Eyssel, F., Zou, J. & Schiebinger, L. Sex and gender analysis improves science and engineering. *Nature* **575**, 137–146 (2019).

77. Brooke, S. Trouble in programmer's paradise: gender-biases in sharing and recognising technical knowledge on stack overflow. *Information, Communication & Society* 1–22 (2021).

78. Altmetric. How it works. https://www.altmetric.com/about-our-data/how-it-works-2/ (2021).

79. Peng, H., Romero, D. M. & Horvát, E.-Á. Dynamics of cross-platform attention to retracted papers. *PNAS* (2022).

80. Elsevier. All science journal classification codes. http://bartoc.org/en/node/20290 (2021).

81. Sinha, A. *et al.* An overview of microsoft academic service (mas) and applications. In *WWW* (2015).

82. Microsoft. How microsoft academic uses knowledge to address the problem of conflation and disambiguation. https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-uses-knowledge-address-problem-conflation-disambiguat (2018).

83. gendercomputer. https://github.com/tue-mdse/genderComputer.

84. Genderize. https://genderize.io/ (2021).

85. Gender detection. http://codingnews.info/post/gender-detection.html (2015).

86. Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M. & Strohmaier, M. Inferring gender from names on the web: A comparative evaluation of gender detection methods. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 53–54 (2016).

87. Jia, J. & Zhao, Q. Gender prediction based on chinese name. In *CCF International Conference on Natural Language Processing and Chinese Computing*, 676–683 (Springer, 2019).

88. Santamaría, L. & Mihaljević, H. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science* **4**, e156 (2018).

89. Torvik, V. I. & Agarwal, S. Ethnea–an instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database. *International Symposium on Science of Science* (2016).

90. fuzzywuzzy. https://github.com/seatgeek/fuzzywuzzy.

91. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* (2015).

92. Banshal, S. K., Singh, V. K., Muhuri, P. K. & Mayr, P. Disciplinary variations in altmetric coverage of scholarly articles. *arXiv:1910.04205* (2019).

Acknowledgements We thank Altmetric and Microsoft Academic Graph for sharing the data used in this study. We also thank IOP Publishing for sharing their submission data used in our gender validation. The

authors thank Orsolya Vásárhelyi and Bogdan Vasilescu for helpful discussion. This work has been partially funded by NSF CAREER Grant No IIS-1943506 and by the Air Force Office of Scientific Research under award number FA9550-19-1-0029.

Author Contributions H.P., M.T., D.R., and E.Á.H. collectively designed the study; H.P. performed the analyses; H.P., M.T., D.R., and E.Á.H. wrote the manuscript.

Competing Interests The authors declare no competing interests.

Additional Information Supplemental material is available for this paper.

Data availability The Altmetric data can be accessed free of charge by researchers from `https://www.altmetric.com/research-access/`. The Microsoft Academic Graph is publicly available at `https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/` or `https://www.microsoft.com/en-us/research/project/open-academic-graph/`. A public repository of our data will be available at `https://doi.org/xyz/m9.figshare.xyz`.

Materials & Correspondence Correspondence should be addressed to D.M.R. and E.Á.H.

Supplementary Files

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

- si.pdf