



The extent and drivers of gender imbalance in neuroscience reference lists

Jordan D. Dworkin^{ID 1}, Kristin A. Linn^{ID 1}, Erin G. Teich^{ID 2}, Perry Zurn³, Russell T. Shinohara¹ and Danielle S. Bassett^{ID 2,4,5,6,7,8}✉

Similarly to many scientific disciplines, neuroscience has increasingly attempted to confront pervasive gender imbalances. Although publishing and conference participation are often highlighted, recent research has called attention to the prevalence of gender imbalance in citations. Because of the downstream effects of citations on visibility and career advancement, understanding the role of gender in citation practices is vital for addressing scientific inequity. Here, we investigate whether gendered patterns are present in neuroscience citations. Using data from five top neuroscience journals, we find that reference lists tend to include more papers with men as first and last author than would be expected if gender were unrelated to referencing. Importantly, we show that this imbalance is driven largely by the citation practices of men and is increasing over time as the field diversifies. We assess and discuss possible mechanisms and consider how researchers might approach these issues in their own work.

In recent years, science has been pushed to grapple with the vast gender imbalances in academic participation. In addition to large and persistent gaps in the proportion of women across scientific fields¹, research has identified gender imbalances along various measures of academic inclusion and success. Such inequalities have been found in compensation², grant funding^{3,4}, credit for collaborative work⁵, teaching evaluations^{6–8}, hiring and promotions^{9,10}, authorship^{11–13} and citations^{14–16}. Importantly, although this study focuses on gender, similar biases exist in other domains, including race, socioeconomic status and university prestige^{17–19}.

While many aspects of gender bias have yet to be studied within neuroscience specifically, issues of gender and diversity have commanded increasing attention over the past several years. Groups such as BiasWatchNeuro (<http://biaswatchneuro.com/>), Women in Neuroscience (<http://winrepo.org/>) and Anne's List (<http://anneslist.net/>) have been created to track and promote the inclusion of women in conferences and symposia. Furthermore, major neuroscience societies have publicly discussed ways to improve representation²⁰, and journals have sought to balance the composition of editors and reviewers²¹. On the heels of these efforts, a recent study showed that authorship and public speaking have indeed become more balanced in the last decade²².

Despite progress in these areas, the presence of differential engagement with scholarship could lead to prolonged inequities in other areas. Recent studies of such engagement have found not only that people from marginalized groups are undercited in fields such as communications²³ and philosophy²⁴, but also that women-led research in particular tends to be undercited in astronomy¹⁶, international relations¹⁵ and political science²⁵. Theoretical work has proposed a 'Matilda effect' in which the contributions of men are seen as more central and are, therefore, sought out more often and

evaluated more highly²⁶. The presence of such an effect in scientific authorship would likely produce reputational and citational inequity. In this case, women-led work could remain under-discussed and perceived as marginal compared to men-led work.

Because of the potential downstream effects of inequitable engagement with women- and men-led work, the study of citation behavior is a critical endeavor for understanding and addressing biases in a particular field. Additionally, achieving gender equity within reference lists can be pursued by all researchers (unlike, for example, achieving gender equity within keynote-speaker roles). Thus, in this study we seek to determine the existence and potential drivers of gender imbalance in neuroscience citations. Previous work in citation gaps has often focused on citation counts of papers^{15,16}, finding that work by women tends to receive fewer citations than similar work by men. Yet this formulation only measures the passive consequences of gendered citation behavior, rather than directly measuring the behavior itself. Instead, building on recent studies in international relations and political science^{25,27}, we investigate the relationship between authors' gender and the gender makeup of their reference lists. Using this framework, we are able to quantify properties associated with authors serving as both objects and agents of undercitation.

For this study, we examine articles published in five top neuroscience journals since 1995. Within this pool of articles, we obtain probabilistic estimates of author gender, find connections between citing and cited papers, locate and remove self-citations and study the links between the gender of authors and their role as objects and agents of undercitation. Specifically, we test the following hypotheses: (1) the overall citation rate of women-led papers (defined here as papers with women as the first and/or last author) will be lower than expected given the relevant characteristics

¹Penn Statistics in Imaging and Visualization Center, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ²Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Philosophy and Religion, American University, Washington, DC, USA. ⁴Department of Physics & Astronomy, College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA. ⁵Department of Electrical & Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA. ⁶Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁷Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁸Santa Fe Institute, Santa Fe, NM, USA.

✉e-mail: dsb@seas.upenn.edu

of papers; (2) undercitation of women-led papers will occur to a greater extent within men-led reference lists; (3) undercitation of women-led papers will be decreasing over time but at a slower rate within men-led reference lists; and (4) differences in undercitation between men-led and women-led reference lists will be partly explained by the structure of authors' social networks. Significance was assessed for these hypotheses using a citation graph-preserving null model, and all P values were corrected for multiple comparisons.

Results

Data description. We extracted data from the Thomson Reuters' Web of Science (WoS) database for research articles and reviews published in five top neuroscience journals since 1995. We selected *Nature Neuroscience*, *Neuron*, *Brain*, *Journal of Neuroscience* and *Neuroimage*, as these journals were reported by the WoS to have the highest Eigenfactor scores²⁸ in the neuroscience category. In all, 61,416 articles were included in the dataset of citing and cited papers. Details of the procedures used to obtain and disambiguate full author names can be found in the Methods.

Gender was assigned to authors' first names using a combination of two publicly available probabilistic databases (Methods). The label 'woman' was assigned to authors whose name had a probability ≥ 0.70 of belonging to someone identifying as a woman; likewise, the label 'man' was assigned to authors whose name had a probability ≥ 0.70 of belonging to someone identifying as a man. In a random sample of 200 authors, the accuracy of these automated assignments was 0.96 (Supplementary Text and Supplementary Tables 1 and 2). We performed the following analyses using the articles for which gender could be assigned with high probability to the first and last author (88%; $n=54,225$), but sensitivity analyses were also conducted on the full data (Supplementary Tables 3 and 4).

Given the limitations of probabilistic analyses, the authors may in fact have a sex or gender different from the one we have assigned and/or be intersex, transgender or nonbinary^{29,30}. In some cases, citers will know the sex and/or gender of the authors they cite. In many cases, they will not know, but rather infer the gender of the authors they cite. Instances of both known and inferred gender have the potential to incite either explicit or implicit bias in citing authors^{6,31,32}. Our probabilistic analysis by gendered name, therefore, functions to nontrivially capture bias arising due to both known and inferred gender in citation practices.

Trends in authorship. Across the articles in the sample, the proportion of articles with a woman as first or last author significantly increased between 1995 and 2018, at a rate of approximately 0.60% each year (95% confidence interval (CI) = (0.53%, 0.67%)). This trend varied across journals, with the *Journal of Neuroscience* (0.67%; 95% CI = (0.57%, 0.77%)), *Neuroimage* (0.89%; 95% CI = (0.72%, 1.06%)) and *Brain* (1.16%; 95% CI = (0.92%, 1.41%)) all showing increases of between 0.65% and 1.2% per year. *Neuron* showed a modest increase of 0.29% per year (95% CI = (0.12%, 0.45%)), but *Nature Neuroscience* did not show a clear increasing trend (0.19%; 95% CI = (-0.09%, 0.46%)). Across these five journals, the overall proportion of articles that had women as first or last author increased from 36% in 1995 to 50% in 2018 (Fig. 1).

Citation imbalance relative to overall authorship proportions. To quantify citation behavior within neuroscience articles, we specifically examined the reference lists of papers published between 2009 and 2018 ($n=31,418$). Thus, while all papers in the dataset were potential cited papers, references to citing papers refer only to those published since 2009. For each citing paper, we took the subset of its citations that had been published in one of the above five journals since 1995 and determined the gender of the cited first and last authors. We removed self-citations (defined as cited papers for which either the first or last author of the citing paper matched the

first or last author of the cited paper; see Methods) from consideration for all analyses presented in the main text. See Supplementary Fig. 1 and Supplementary Tables 5–8 for detailed analyses of the role of self-citations. We then calculated the number of cited papers that fell into each of the four first author and last author categories: man and man (MM), woman and man (WM), man and woman (MW) and woman and woman (WW). Single-author papers by men and women were included in the MM and WW categories, respectively.

As a simple first step, we compared the observed number of citations within each category to the number that would be expected if references were drawn randomly from the pool of papers (Fig. 2a). To obtain the number that would be expected under this assumption of random draws, we calculated the gender proportions among all papers published previous to the citing paper, thus representing the proportion among the pool of papers that the authors could have cited, and multiplied them by the number of papers cited. The following section expands this naive measure to account for potential relationships between author gender and other relevant characteristics of cited papers.

Of the 303,886 citations given between 2009 and 2018, MM papers received 61.7%, compared to 23.6% for WM papers, 9.0% for MW papers and 5.8% for WW papers. The expected proportions based on the pool of citable papers were 55.3% for MM, 26.2% for WM, 10.2% for MW and 8.3% for WW. We defined a measure of over/undercitation as the (percentage observed – percentage expected)/percentage expected. This measure thus represents the percentage over/undercitation relative to the expected proportion (see Methods for further details and Supplementary Table 9 for a sensitivity analysis using article-level over/undercitation). By this measure, MM papers were cited 11.6% more than expected (95% CI = (11.2%, 12.0%)), WM papers were cited 10.1% less than expected (95% CI = (-10.7%, -9.5%)), MW papers were cited 12.5% less than expected (95% CI = (-13.6%, -11.4%)) and WW papers were cited 30.2% less than expected (95% CI = (-31.3%, -29.0%)). These values correspond to MM papers being cited roughly 19,500 more times than expected, compared to 8,000 fewer times for WM papers, 3,900 fewer times for MW papers and 7,600 fewer times for WW papers.

Citation imbalance after accounting for relevant characteristics of papers. The comparison of citations to overall authorship proportions does not take into account other properties of published papers that may make them more or less likely to be cited by later scholarship. The potential relationships between author gender and other characteristics of papers make it difficult to isolate links between gender and citation rates. To address this issue, we sought to model, and then account for, any relationships between gender and paper characteristics. We selected five features of a paper as being potentially relevant for citation: (1) year of publication, (2) journal of publication, (3) number of authors, (4) research article or review and (5) first and last author seniority. We then modeled the multinomial gender category (that is, MM/WM/MW/WW) as a function of the above characteristics using a generalized additive model (GAM), yielding estimated gender probabilities for each paper. See Methods for details and Supplementary Fig. 2 for the impact of alternate specifications.

We next sought to compare the observed citation rates to those that would be expected if only these non-gender characteristics were relevant to citation. To do so, we compared the true gender category of each cited paper to its estimated probabilities of belonging to each of the four categories. In contrast to the previous section, in which gender probabilities model citation as a random draw from the existing literature, these probabilities can be viewed as the expected gender proportions across random draws from a narrow pool of papers highly similar to the cited paper (Fig. 2b).

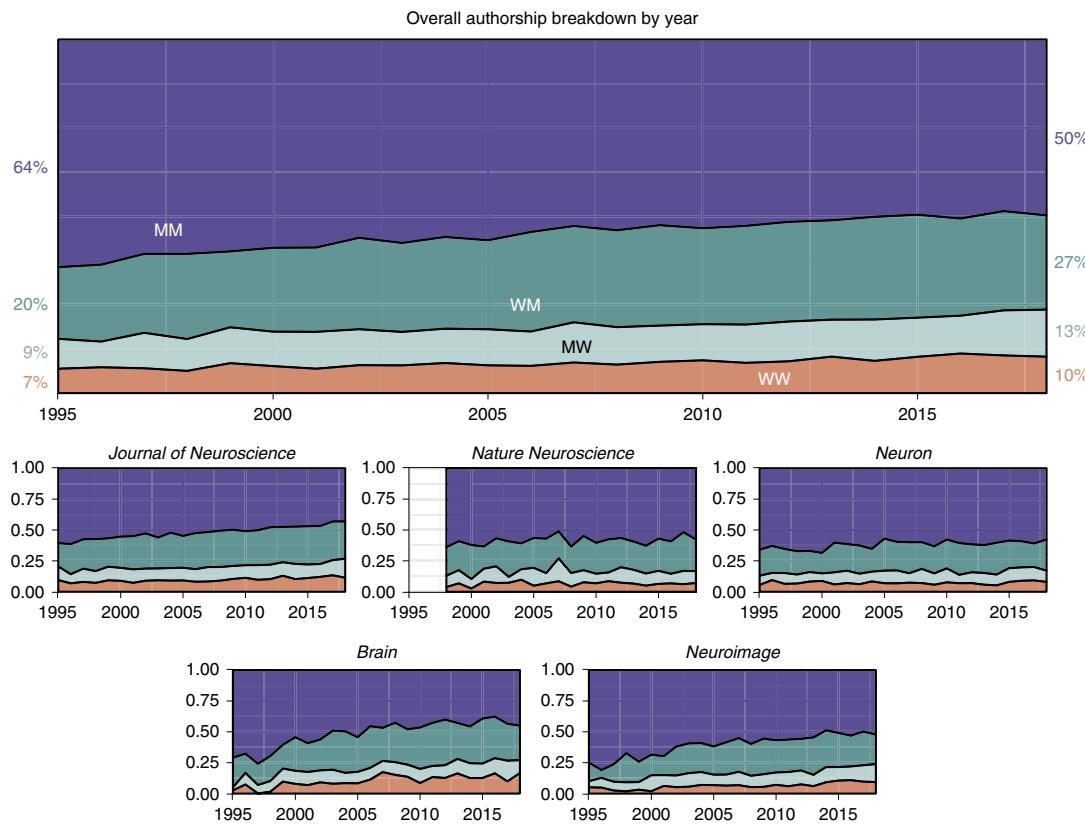


Fig. 1 | Trends in author gender within top neuroscience journals between 1995 and 2018. Overall trends across the five journals studied (top) and trends within each journal (bottom); proportions are shown of articles with men as first and last author (purple), women as first author and men as last author (green), men as first author and women as last author (gray) and women as both first and last author (orange). Note that *Nature Neuroscience* was established in 1998. MM, man first author and man last author; WM, woman first author and man last author; MW, man first author and woman last author; WW, woman first author and woman last author.

These probabilities can also be viewed as the expected proportions if the gender of the authors of a cited paper were randomly swapped across highly similar papers. This second framing makes clear that the following results do not depend on breaking the structure of the citation graph; it is also the basis of the graph-preserving null model used to assess statistical significance (Methods, Supplementary Fig. 3 and Supplementary Table 10).

Summing up the number of cited papers from each category again gives us the observed citation rates, and summing up the estimated gender probabilities across all cited papers gives us new expected citation rates. As reported above, the observed citation rates were 61.7% (MM), 23.6% (WM), 9.0% (MW) and 5.8% (WW) across gender categories. Based on the relevant properties of cited papers, the new expected citation rates were 58.6% for MM, 25.3% for WM, 9.4% for MW and 6.7% for WW. Thus, after accounting for salient non-gender characteristics, MM papers were still cited 5.2% more than expected (95% CI = (4.8%, 5.5%), $P < 0.001$), WM papers were cited 6.7% less than expected (95% CI = (-7.3%, -6.0%), $P = 0.008$), MW papers were cited 4.6% less than expected (95% CI = (-5.7%, -3.3%), $P = 0.86$) and WW papers were cited 13.9% less than expected (95% CI = (-15.2%, -12.5%), $P = 0.003$). Of 303,886 total citations, these values correspond to citations being given to MM papers around 9,300 more times than expected, compared to approximately 5,100 fewer times for WM papers, 1,300 fewer times for MW papers and 2,800 fewer times for WW papers. The observed overcitation of MM papers and undercitation of WM and WW papers provide support for the hypothesis that the citation rate of women-led papers is lower than expected given relevant characteristics (hypothesis 1).

The effect of author gender on citation behavior. By focusing the present analyses on the gender makeup of reference lists, as opposed to the number of citations that articles receive, we were able to investigate the gender of the citing authors in addition to that of the cited authors. Thus, in this section we compare the gender makeup of references within papers that had men as both first and last author (referred to again as MM) to those within papers that had women as either first or last author (henceforth referred to as WUW, comprising WM, MW and WW papers). Of the 31,418 articles published in one of the five journals between 2009 and 2018, roughly 51% were MM and 49% were WUW.

After separating citing articles by author gender, we found that the imbalance within reference lists shown previously was driven largely by the citation practices of MM teams. Specifically, within MM reference lists ($n = 146,077$ citations), other MM papers were cited 8.0% more than expected (95% CI = (7.6%, 8.5%), $P < 0.001$), WM papers were cited 9.3% less than expected (95% CI = (-10.2%, -8.3%), $P < 0.001$), MW papers were cited 9.0% less than expected (95% CI = (-10.6%, -7.2%), $P = 0.10$) and WW papers were cited 23.4% less than expected (95% CI = (-25.5%, -21.7%), $P < 0.001$; Fig. 3a). Within WUW reference lists ($n = 130,953$ citations), MM papers were cited only 2.5% more than expected (95% CI = (2.0%, 3.1%), $P = 0.07$), WM papers were cited 4.6% less than expected (95% CI = (-5.7%, -3.7%), $P = 0.12$), MW papers were cited 0.1% less than expected (95% CI = (-2.0%, 1.8%), $P > 0.99$) and WW papers were cited 4.2% less than expected (95% CI = (-6.5%, -1.9%), $P > 0.99$; Fig. 3a). The observed differences between MM and WUW reference lists were all significant ($P < 0.0001$), providing support for the

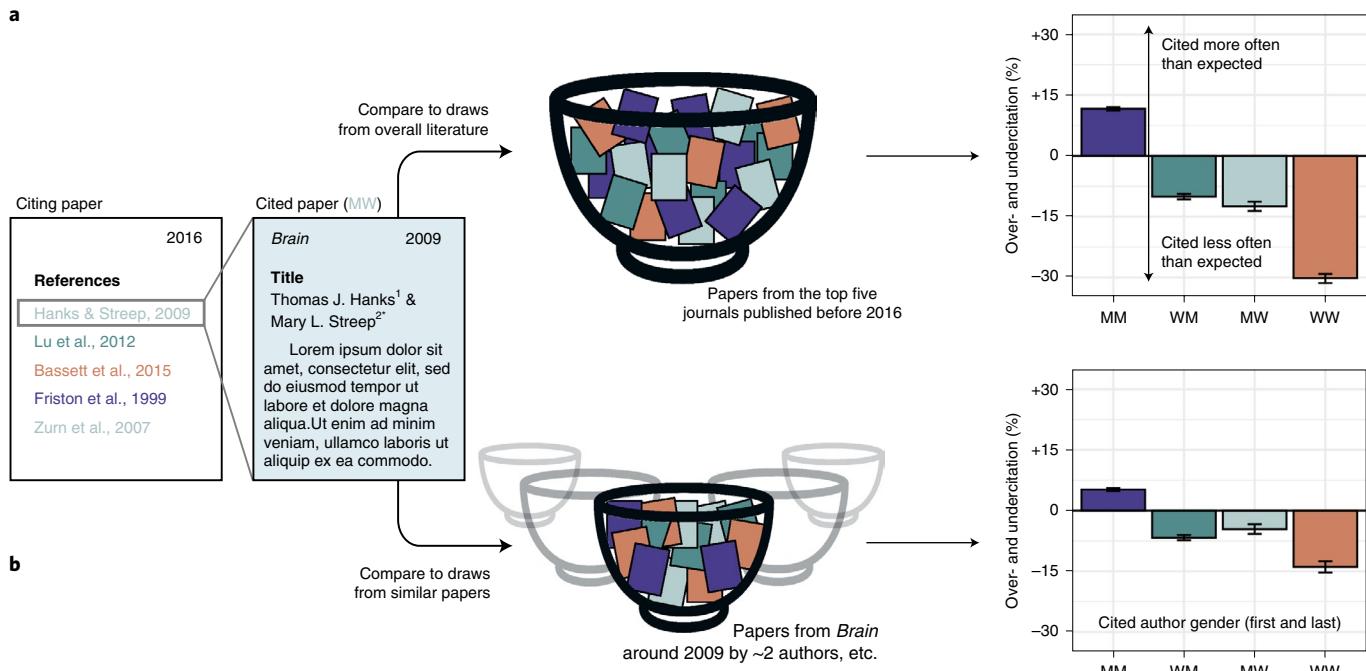


Fig. 2 | Construction and visualization of overall over- and undercitation measures. a, Illustration of the random-draws model, in which gender proportions in reference lists are compared to the overall gender proportions of the existing literature. The percentage of over- and undercitation of different author gender groups is compared to their expected proportions under the random-draws model (right). **b,** Illustration of the relevant characteristics model, in which gender proportions in reference lists are compared to gender proportions of articles that are similar to those that were cited across various domains. Over- and undercitation of different author gender groups is shown compared to their expected proportions under the relevant characteristics model (right). Bars represent overall over- and undercitation according to a given model, calculated from 303,886 total citations. Error bars represent the 95% CI of each over- and undercitation estimate, calculated from 1,000 bootstrap resampling iterations.

hypothesis that undercitation of women-led papers occurs to a greater extent within men-led reference lists (hypothesis 2).

Within the WUW group, the citation proportions of the WM, MW and WW subgroups suggest a more fine-grained link between the increased citation of women-led work and the increased leadership role of women on the citing team (Fig. 3b). Specifically, WM teams still slightly undercite WW papers relative to expectation, but do so at approximately half the rate of MM teams. MW reference lists contain roughly the expected citation proportions across gender groups, and WW reference lists contain slightly more MW and WW papers than expected (overciting WW papers at around half the rate that MM teams undercite WW papers).

This moderate overcitation of women-led work within women-led reference lists points to a possible role of social networks in shaping citation behavior, a possibility that we explore in detail later. Additional supplementary analyses were performed to determine the potential role of alternative mechanisms, such as research subfields, highly cited papers and seniority (Supplementary Figs. 4–6). These factors were found to have little impact on either the extent of citation imbalance or the gender differences in citation behavior.

Temporal trends of citation imbalance. In addition to overall citation behavior, we sought to quantify time-varying gender imbalance as the field has become more diverse. As an intuitive univariate measure of citation imbalance, we examined the yearly absolute difference between the observed and expected proportion of MM citations. We found that the gap between observed and expected proportions has been growing at a rate of around 0.41 percentage points per year (95% CI = (0.34, 0.49), $P < 0.001$). Importantly, this growing gap does not simply reflect the propensity of authors to cite older literature from when the field was more

dominated by men, as the expected proportions account for the publication year of the articles being cited. This finding suggests that citation practices are becoming less reflective of an increasingly diverse body of researchers, thereby standing in contrast to the hypothesis that undercitation of women-led papers will be decreasing over time (hypothesis 3).

Upon splitting by gender of the citing author, we found that the degree of overcitation has been increasing faster within MM reference lists than within WUW reference lists. Specifically, the absolute difference between the observed and expected proportions of MM citations is growing at a rate of 0.54 percentage points per year (95% CI = (0.43, 0.63), $P < 0.001$; Fig. 4a) within MM reference lists, compared to a rate of 0.29 percentage points per year (95% CI = (0.17, 0.40), $P = 0.023$; Fig. 4a) within WUW reference lists. The fact that overcitation of MM papers is rising faster in MM reference lists than in WUW reference lists ($P = 0.014$) is related to the second aspect of hypothesis 3, although the predicted temporal trend is flipped.

Further analysis revealed that the increasing overcitation of men reflects relatively stable citation proportions for MM papers in the face of decreasing expected proportions over time (Fig. 4b). In fact, the observed proportion of MM citations has been increasing slightly within MM reference lists, at a rate of approximately 0.15 percentage points per year (95% CI = (0.03, 0.26)). This proportion has not been clearly increasing or decreasing within WUW reference lists, changing with a rate of -0.08 percentage points per year (95% CI = (-0.19, 0.04)). These findings demonstrate that, although the rate at which scholars cite men has been relatively stable, this lack of change has led the gender proportions within reference lists to be increasingly unrepresentative of the diversifying field.

The relationship between social networks and citation behavior. Recent work has shown that researchers are more likely to work

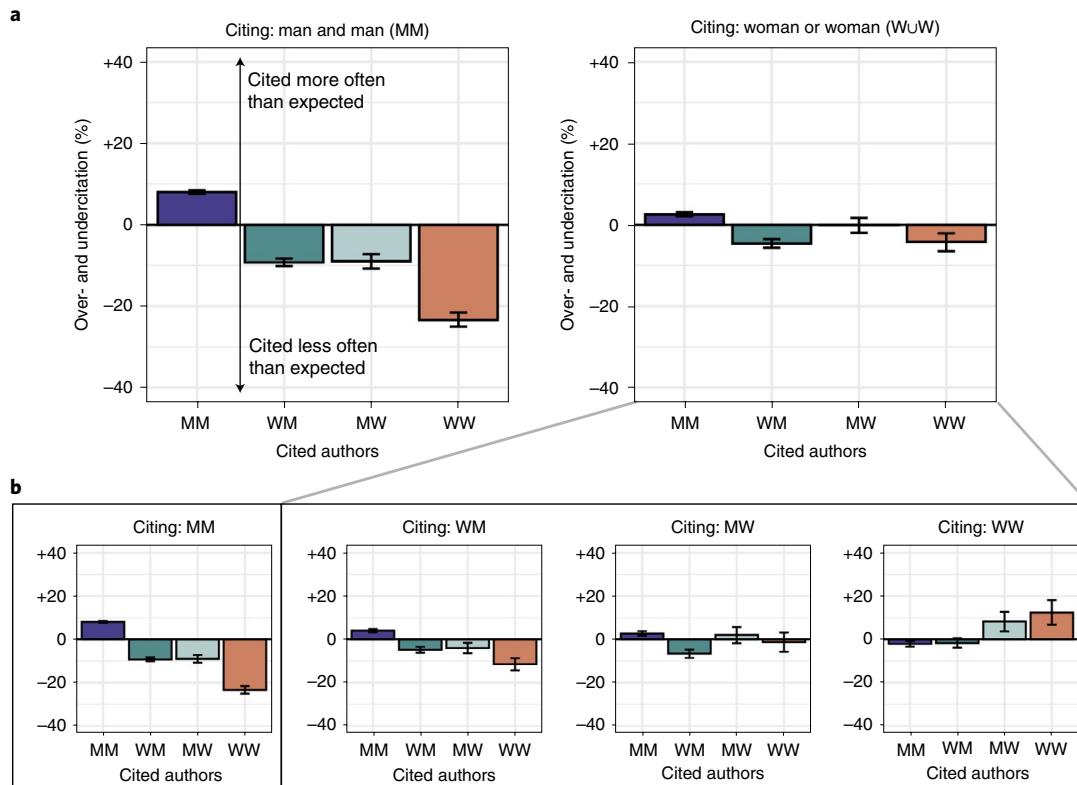


Fig. 3 | Relationship between author gender and gendered citation practices. **a**, Degree of over- and undercitation of different author genders within MM and WUW reference lists. Papers with men as both first and last author overcite men to a greater extent than papers with women as either first or last author. **b**, Full breakdown of gendered citation behavior within MM, WM, MW and WW reference lists. Bars represent overall over/undercitation, calculated from 277,030 total citations (146,077 within MM reference lists and 130,953 within WUW reference lists). Error bars represent the 95% CI of each over- and undercitation estimate, calculated from 1,000 bootstrap resampling iterations.

with other researchers of their own gender (that is, homophily exists within co-authorship networks), and that local homophily in social networks can affect perceptions of the overall network^{33,34}. Because perception and affinity biases could be potential drivers of the overcitation of men by men and slight overcitation of women by women, we sought to estimate the relationship between authors' social networks and their citation behavior. We developed two measures to quantify gender imbalance in the co-authorship network of a paper's authors. Importantly, two papers written by the same authors may have different values for these measures, because the co-authorship network surrounding the authors may change over time.

For a given paper p , we defined man-author over-representation as the difference between (1) the proportion of men within p 's author neighborhood (the union of researchers who previously co-authored a paper with either the first or last author of p) and (2) the overall proportion of men within the network at the time of p 's publication. We additionally defined MM-paper over-representation as the difference between (1) the proportion of MM papers within p 's paper neighborhood (the union of papers written by any previous co-authors of p 's first or last author) and (2) the overall proportion of MM papers within the network at the time of p 's publication. Visual examples of these two measures can be seen in Fig. 5a,b (see Methods for further details).

Co-authorship networks tended to include more men than the base rate in the field, but this feature was especially pronounced within the networks of men-led teams. The median MM team had around 8.2% more men in their co-authorship network than the field's base rate (95% CI=(8.0%, 8.4%); $n=14,151$ MM teams), compared to the median WW team, which had around 3.8% more men (95% CI=(3.3%, 4.5%); $n=2,668$ WW teams). Networks of

mixed-gender teams contained around 6% more men than the field's base rate (WM=6.4%, 95% CI=(6.1%, 6.7%), $n=7,711$ WM teams; MW=5.7%, 95% CI=(5.2%, 6.1%), $n=3,106$ MW teams; Fig. 5c). Local over-representation of MM papers also differed based on author gender. In this case, MM papers were over-represented relative to the field's base rate only within the networks of MM teams (+4.2%, 95% CI=(4.1%, 4.4%)) and WM teams (+2.5%, 95% CI=(2.3%, 2.8%)). MM papers were roughly proportionally represented within the networks of MW teams (+0.7%, 95% CI=(0.2%, 1.0%)) and slightly under-represented within networks of WW teams (-0.4%, 95% CI=(-0.8%, -0.1%); Fig. 5c).

Because gendered differences in social networks followed similar patterns to gendered differences in citation behavior, we sought to determine whether the composition of authors' social networks accounts for overcitation of men. Here, we again utilized the absolute difference between the observed proportion of MM citations within a paper's reference list and the expected proportion based on the characteristics of the cited papers. Without accounting for differences in authors' social networks, we found that the median MM team overcites MM papers by around 5.5 percentage points (95% CI=(5.1, 5.8), $P<0.001$), compared to 3.0 for WM teams (95% CI=(2.6, 3.6), $P<0.001$), 2.4 for MW teams (95% CI=(1.6, 2.9), $P=0.008$) and -0.7 for WW teams (95% CI=(-1.7, 0.3), $P>0.99$; Fig. 6a,c).

To estimate and account for the role of authors' social networks, we modeled the MM overcitation of papers as a function of author gender, man-author over-representation (MA_{or}) and MM-paper over-representation (MMP_{or}). Because the overcitation measure is bounded and skewed, we performed quantile regression to obtain estimates of the conditional median (see Methods for further details and Supplementary Table 11 for an alternative specification). Both

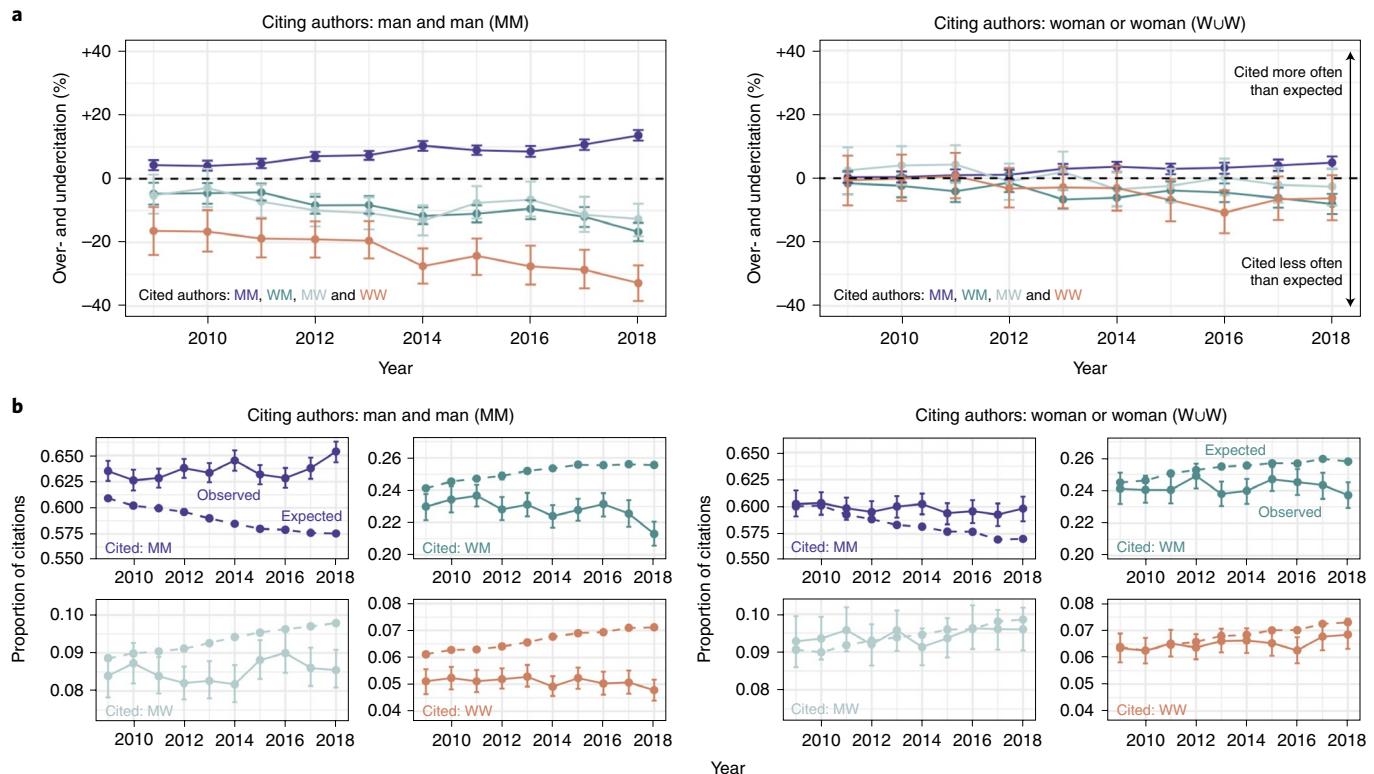


Fig. 4 | Temporal trends in citation rates across gender of the cited and citing author. **a**, Extent of over/undercitation across author gender categories over time, within MM and WUW reference lists. Points represent over/undercitation within the literature in a given year, calculated from 277,030 total citations (146,077 within MM reference lists and 130,953 within WUW reference lists). Error bars represent the 95% CI of each over/undercitation estimate, calculated from 1,000 bootstrap resampling iterations. **b**, Observed (solid line) and expected (dashed line) citation proportions within MM and WUW reference lists. Within each section, clockwise from top left, is shown the observed and expected proportion of citations given to MM papers, WM papers, MW papers and WW papers over time. The figure demonstrates relatively static observed proportions across groups, while expected proportions change due to increasing diversity within the field. The points represent observed or expected citation rates within the literature in a given year. Error bars represent the 95% CI of the observed citation rate, calculated from 1,000 bootstrap resampling iterations.

MMP_{or} and MA_{or} were independently associated with MM overcitation; a one-percentage-point increase in local over-representation of MM papers corresponded to a 0.24-percentage-point increase in median MM overcitation (95% CI = (0.21, 0.28), $P < 0.001$), and a one-percentage-point increase in local over-representation of man authors corresponded to a 0.09-percentage-point increase in median MM overcitation (95% CI = (0.05, 0.12), $P < 0.001$). These relationships were consistent after accounting for author seniority, although they appear to be slightly stronger among more senior teams (Supplementary Fig. 6). Thus, the data do support the hypothesis that a relationship exists between local co-authorship networks and citation behavior (hypothesis 4).

However, after accounting for the structure of authors' social networks, gendered citation patterns remained. Specifically, conditional on authors' networks being gender balanced (that is, MA_{or} = 0 and MMP_{or} = 0), the median MM team would still be expected to overcite MM papers by around 3.5 percentage points (95% CI = (3.1, 3.9), $P < 0.001$), compared to 1.9 for WM teams (95% CI = (1.4, 2.5), $P = 0.023$), 1.6 for MW teams (95% CI = (0.7, 2.3), $P = 0.18$) and -0.4 for WW teams (95% CI = (-1.0, 0.7), $P > 0.99$; Fig. 6b,d). These results suggest that approximately two-thirds of the observed overcitation of men by other men remains after accounting for social networks, while women-led teams tend to cite at nearly proportional rates.

Discussion

As in many scientific disciplines, the field of neuroscience currently faces many structural and social inequities, including marked gender

imbalances²². Although the task of addressing these imbalances often depends on people in positions of power (for example, journal editors²¹, grant reviewers and agencies^{3,4}, department chairs^{9,10} and presidents of scientific societies²⁰), many imbalances are caused and perpetuated by researchers at all levels. One example is imbalance within citation practices^{14,15}. Although the usefulness of citations as a measure of scientific value is tenuous³⁵, the engagement that they represent can affect how central to a field scholars are viewed to be by their peers¹⁴, impacting speaking invitations, grants, awards, tenure, promotion, inclusion in syllabi and teaching evaluations.

In this study, we sought to determine whether there is evidence of gender imbalance in neuroscience citations. We indeed find that neuroscience reference lists tend to include more papers with men as first and last author than would be expected if gender were not a factor. Importantly, this undercitation of women remains after accounting for relevant characteristics of papers and is driven largely by the citation practices of men. Specifically, papers with men as first and last author overcite other MM papers by 8% relative to the expected proportion, undercite WM papers by 9%, undercite MW papers by 9% and undercite WW papers by 23%. For papers with women in one or both primary authorship positions, these values are +2.5%, -4.5%, -0.1% and -4%, respectively. These findings are consistent with results from other fields showing that men are less likely to cite work by women^{14,15,25}.

Gender inequity is understood to result from both systemic and individual bias. Systemic bias refers to discriminatory values, practices and mechanisms that function at the intergroup level in the domain

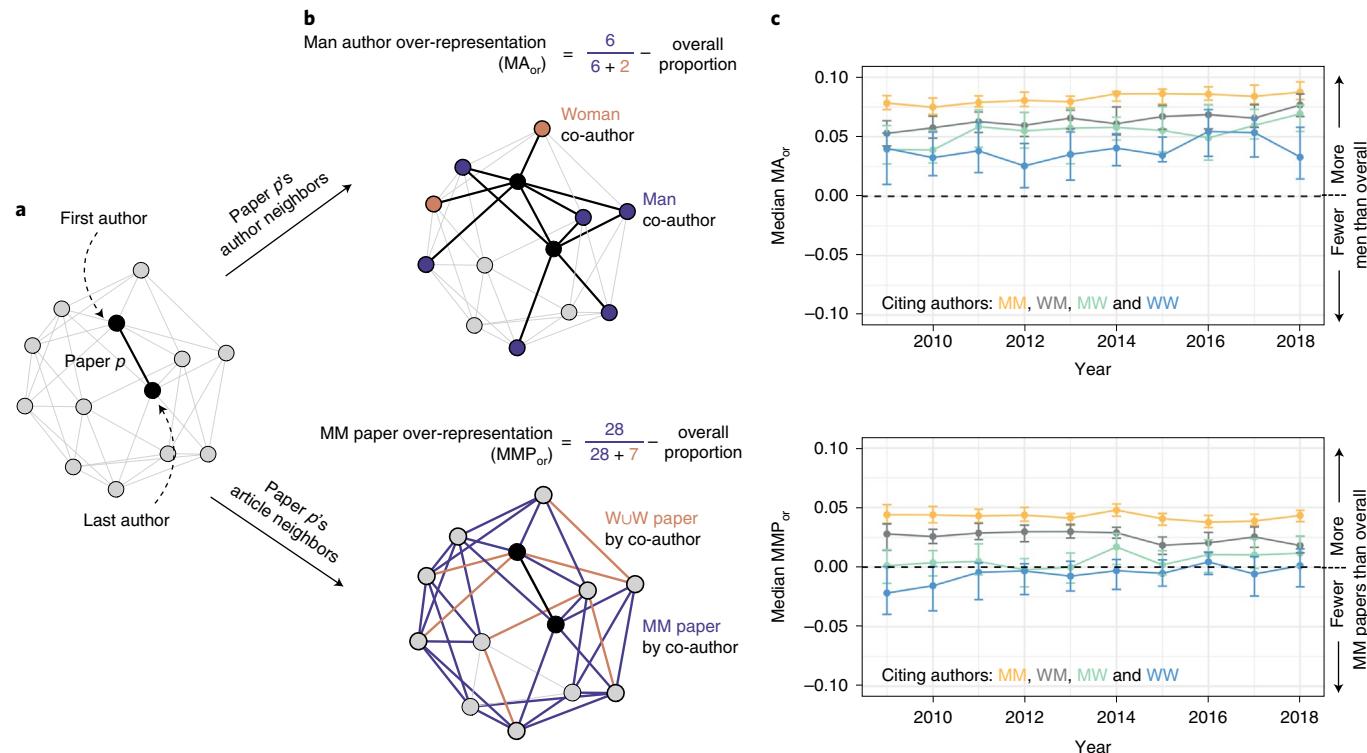


Fig. 5 | Visualization of co-authorship network composition measures. **a**, Example region of a co-authorship network, where a specific article (edge) and the first and last authors (nodes) are highlighted. **b**, Examples of the calculation of man-author over-representation (MA_{or}) and MM-paper over-representation (MMP_{or}) for the highlighted article. Here, MA_{or} is the difference between the local proportion of men (purple nodes) and the overall proportion of men. The quantity MMP_{or} is the difference between the local proportion of MM papers (purple edges) and the overall proportion of MM papers. **c**, Differences in the local network composition based on author gender. Papers with more women tend to have less over-representation of men and men-led papers within their local networks. The points represent man-author and MM-paper over-representation among authors in a given year, calculated from 27,636 author teams ($n=14,151$ (MM), $n=7,711$ (WM), $n=3,106$ (MW) and $n=2,668$ (WW)). Error bars represent the 95% CI of each over-representation estimate, calculated from 1,000 bootstrap resampling iterations.

of social institutions³⁶. Bias at the level of individuals may be either explicit, consisting of consciously held or expressed prejudice against a particular group³⁷, or implicit, consisting of subconsciously harbored discriminatory attitudes that can result in prejudicial speech and social behaviors³². Implicit bias with respect to names has previously been shown in studies of race-based^{31,38} and gender^{6,39} discrimination. The undercitation of women in neuroscience papers, therefore, may be due to systemic and/or individual gender bias, relative to either the known gender of an author or an author's gendered name.

In seeking to understand the drivers of gendered citation patterns, we find that imbalances within the social networks of authors account for roughly one-third of the observed overcitation of men. Other considerations, such as research subfields and high-impact papers, appear to have only a small role. Of the imbalances that remain, the marked difference in citation behavior between men-led and women-led teams is of particular interest. One potential explanation for this behavior gap is greater conscious or unconscious bias among men, which could induce more negative evaluations of women-led work. This explanation would be consistent with studies showing evaluative bias in graduate admissions⁴⁰, faculty hiring², grant funding⁴ and promotion¹⁰.

Upstream imbalances, such as the over-representation of men in course syllabi⁴¹ and conference speaking roles²², could also plausibly explain citation imbalance. Yet these mechanisms would likely yield an overall overcitation of men that, unlike the data suggest, does not differ based on the gender of the citing authors. In this case, our observation that women-led teams display less gender citation imbalance could possibly be explained by their conscious efforts to

seek out and cite work by other women. In contrast, indifference or lack of awareness among men could lead to the propagation of upstream imbalances.

Greater awareness of existing, persisting and even increasing imbalances in citation practices is an important step in heightening the willingness of researchers to address these issues. Recent work has offered guidelines for responsible citation practices that consider gender balance⁴². Tools exist to probabilistically measure the proportion of women and men within course syllabi and reference lists⁴³. Various databases can also be used to create representative reference lists. For example, BiasWatchNeuro (<https://biaswatch-neuro.co/>) publishes base rates across neuroscience subfields and Women in Neuroscience (<http://winrepo.org/>) and Anne's List (<http://annestable.net/>) contain detailed, searchable entries of women in neuroscience and their areas of expertise.

Addressing the identified imbalances will require researchers, particularly men, to make use of available resources and engage in more thoughtful citation practices. Efforts can also be made by journal editors and reviewers to inform authors of these issues and encourage transparency within manuscripts. This paper, for example, includes a citation gender diversity statement that describes the gender makeup of its reference list. Educating graduate students about citation practices will also be vital, and such discussions could potentially be included in the Responsible Conduct of Research (RCR) requirements from the National Institutes of Health (NIH) and National Science Foundation (NSF).

Beyond thoughtfulness and ad hoc efforts for achieving gender balance, the ethics of citation practices remain to be defined.

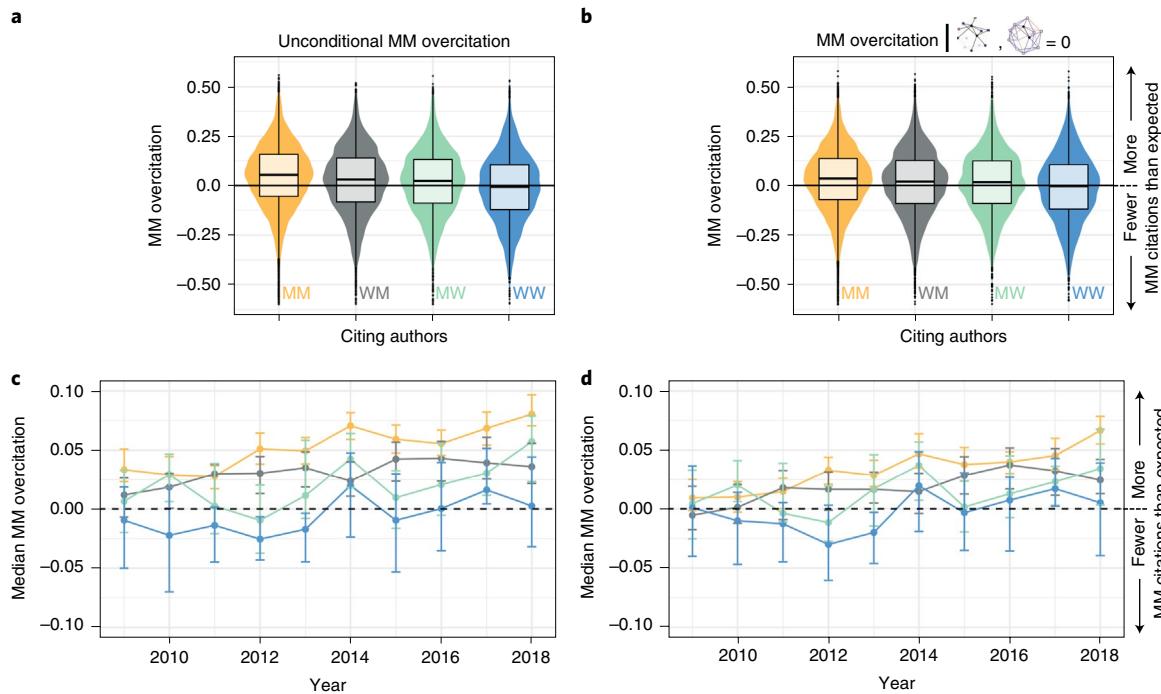


Fig. 6 | Article-level overcitation of MM papers before and after accounting for local network composition. **a**, Overcitation of MM papers by citing author gender. MM, WM and MW papers tend to overcite MM papers relative to expectation, while WW papers cite MM and WW papers at roughly the expected rate. **b**, Overcitation of MM papers by author gender, after accounting for network effects. Local network composition explains some of the group differences, but the general pattern remains. Box and violin plots show the distribution of article-level MM overcitation across 27,636 author teams ($n=14,151$ (MM), $n=7,711$ (WM), $n=3,106$ (MW) and $n=2,668$ (WW)). The center line represents the median, boxes represent the 25th and 75th percentiles, with their length giving the interquartile range (IQR), whiskers represent observations up to 1.5 times the IQR away from the 25th and 75th percentiles, and dots represent outlying observations. **c,d**, Overcitation of MM papers is increasing over time across groups (**c**), even after accounting for the effects of authors' social networks (**d**). Points represent MM overcitation among authors in a given year, calculated from the same 27,636 author teams described above. Error bars represent the 95% CI of each overcitation estimate, calculated from 1,000 bootstrap resampling iterations.

Righting social inequities may be accomplished on a number of different models. On the distributive model, social goods and resources, or, in this case, citations, ought to be allocated according to some morally proper distribution (be it full equality or equity conditional on select features^{44,45}). This paradigm, however, is limited in that it does not seek to address histories and structures of inequality^{46,47}. Diversity models, by contrast, recommend acts of reparative justice⁴⁸, which might include affirmative action⁴⁹ in citational practices or institutional reform to support citation parity.

Distributive and diversity-based models raise important questions for citation ethics. Should gender balance in citation practices reflect random distributions or distributions tuned to relevant features (and, if so, which features)? Are such distributional structures sufficient to either correct for a history of under-representation or secure a future of equitable representation? Given the slow pace of social change and the increasing citation imbalance in neuroscience, is it justifiable for some researchers to overcite papers produced by women-led teams wherever possible? Might the effects of systemic bias be counteracted by the substantial employment of women in decision-making bodies, reforming checkpoints and professional activism? And, given the function of implicit bias and its capacity for correction via experience, should researchers of all genders commit to collaborating more robustly with women and other gender minorities?

Overall, the work of citation is an important element in the research ethics of any field. Insofar as citation patterns today have inescapable effects on the future of neuroscience, citational practices in the field warrant more serious attention.

Limitations and future work. This work is subject to several limitations. First, this study considers five top neuroscience journals, thereby reducing the confound of journal prestige but potentially limiting the generalizability of the results. Second, the methods used for gender determination are limited to binary man and woman gender assignments. This study design, therefore, is not well accommodated to intersex, transgender and/or nonbinary identities and incorrectly assumes that all authors can be placed into one of two categories. Ideally, future work will be able to move beyond the gender binary. Third, this study investigates biases solely along gender lines. Future work could examine biases along other lines, for example, race, class, sexuality, disability and citizenship, as well as their intersections with gender⁵⁰. This knowledge could inform the development of a broader database of under-represented scholars, inspired by the American Philosophical Association's UP Directory (<https://updirectory.gear.host/>).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-020-0658-y>.

Received: 10 January 2020; Accepted: 20 May 2020;
Published online: 19 June 2020

References

1. Holman, L., Stuart-Fox, D. & Hauser, C. E. The gender gap in science: how long until women are equally represented? *PLoS Biol.* **16**, e2004956 (2018).

2. Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, J. Science faculty's subtle gender biases favor male students. *Proc. Natl Acad. Sci. USA* **109**, 16474–16479 (2012).
3. Reshma, J. Sex differences in attainment of independent funding by career development awardees. *Ann. Intern. Med.* **151**, 804–811 (2009).
4. van der Lee, R. & Ellemers, N. Gender contributes to personal research funding success in the Netherlands. *Proc. Natl Acad. Sci. USA* **112**, 12349–12353 (2015).
5. Sarsons, H. Recognition for group work: gender differences in academia. *Am. Econ. Rev.* **107**, 141–145 (2017).
6. MacNell, L., Driscoll, A. & Hunt, A. N. What's in a name: exposing gender bias in student ratings of teaching. *Innov. Higher Educ.* **40**, 291–303 (2015).
7. Mengel, F., Sauermann, J. & Zöllitz, U. Gender bias in teaching evaluations. *J. Eur. Econ. Assoc.* **17**, 535–566 (2019).
8. Boring, A. Gender biases in student evaluations of teaching. *J. Public Econ.* **145**, 27–41 (2017).
9. Nielsen, M. W. Limits to meritocracy? Gender in academic recruitment and promotion processes. *Sci. Pub. Pol.* **43**, 386–399 (2016).
10. De Paola, M. & Scoppa, V. Gender discrimination and evaluators' gender: evidence from Italian academia. *Economica* **82**, 162–188 (2015).
11. 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).
12. Wilhelm, I., Conklin, S. L. & Hassoun, N. New data on the representation of women in philosophy journals: 2004–2015. *Int. J. Philos. Stud.* **175**, 1441–1464 (2018).
13. Huang, J., Gates, A. J., Sinatra, R. & Barabasi, A.-L. Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc. Natl Acad. Sci. USA* **117**, 4609–4616 (2020).
14. Ferber, M. A. & Brun, M. The gender gap in citations: does it persist? *Fem. Econ.* **17**, 151–158 (2011).
15. Maliniak, D., Powers, R. & Walter, B. F. The gender citation gap in international relations. *Int. Organ.* **67**, 889–922 (2013).
16. Caplar, N., Tacchella, S. & Birrer, S. Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nat. Astron.* **1**, 0141 (2017).
17. Fang, D., Moy, E., Colburn, L. & Hurley, J. Racial and ethnic disparities in faculty promotion in academic medicine. *JAMA* **284**, 1085–1092 (2000).
18. Petersen, A. M. et al. Reputation and impact in academic careers. *Proc. Natl Acad. Sci. USA* **111**, 15316–15321 (2014).
19. Way, S. F., Morgan, A. C., Larremore, D. B. & Clauset, A. Productivity, prominence and the effects of academic environment. *Proc. Natl Acad. Sci. USA* **116**, 10729–10733 (2019).
20. Joels, M. & Mason, C. A tale of two sexes. *Neuron* **82**, 1196–1199 (2014).
21. Anonymous. Promoting diversity in neuroscience. *Nat. Neurosci.* **21**, 1 (2018).
22. Schrouff, J. et al. Gender bias in (neuro)science: facts, consequences and solutions. *Eur. J. Neurosci.* **50**, 3094–3100 (2019).
23. Chakravarty, P., Kuo, R., Grubbs, V. & McIlwain, C. #CommunicationSoWhite. *J. Commun.* **68**, 254–266 (2018).
24. Thiem, Y., Sealey, K. F., Ferrer, A. E., Trott, A. M. & Kennison, R. *Just Ideas? The Status and Future of Publication Ethics in Philosophy* (Publication Ethics, 2018).
25. Dion, M. L., Sumner, J. L. & Mitchell, S. M. Gendered citation patterns across political science and social science methodology fields. *Polit. Anal.* **26**, 312–327 (2018).
26. Rossiter, M. W. The Matthew Matilda effect in science. *Soc. Stud. Sci.* **23**, 325–341 (1993).
27. Mitchell, S. M., Lange, S. & Brus, H. Gendered citation patterns in international relations journals. *Int. Stud. Perspect.* **14**, 485–492 (2013).
28. Bergstrom, C. T., West, J. D. & Wiseman, M. A. The Eigenfactor metrics. *J. Neurosci.* **28**, 11433–11434 (2008).
29. Feder, E. K. *Making Sense of Intersex: Changing Ethical Perspectives in Biomedicine* (Indiana University Press, 2014).
30. Stryker, S. *Transgender History (Seal Studies)* (Seal Press, 2008).
31. Bertrand, M. & Mullainathan, S. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* **94**, 991–1013 (2004).
32. Brownstein, M. Implicit bias. in *The Stanford Encyclopedia of Philosophy*. Fall 2019 edn. (ed. Zalta, E. N.) <https://plato.stanford.edu/entries/implicit-bias/> (Stanford University, 2019).
33. Holman, L. & Morandin, C. Researchers collaborate with same-gendered colleagues more often than expected across the life sciences. *PLoS ONE* **14**, e0216128 (2019).
34. Lee, E. et al. Homophily and minority-group size explain perception biases in social networks. *Nat. Hum. Behav.* **3**, 1078–1087 (2019).
35. Aksnes, D. W., Langfeldt, L. & Wouters, P. Citations, citation indicators and research quality: an overview of basic concepts and theories. *SAGE Open* **9**, 215824401982957 (2019).
36. Henry, P. J. Institutional bias. in *Handbook of Prejudice, Stereotyping and Discrimination* (eds. Dovidio, J. F. et al.) 426–440 (Sage, 2010).
37. Clarke, J. A. Explicit bias. *Northwest. Univ. Law Rev.* **113**, 505–586 (2018).
38. Conaway, W. & Bethune, S. Implicit bias and first name stereotypes: what are the implications for online instruction? *J. Online Learn.* **19**, 162–178 (2015).
39. Paludi, M. A. & Strayer, L. A. What's in an author's name? Differential evaluations of performance as a function of author's name. *Sex Roles* **12**, 353–361 (1985).
40. Posselt, J. R. *Inside Graduate Admissions* (Harvard University Press, 2016).
41. Colgan, J. Gender bias in international relations graduate education? New evidence from syllabi. *PS Polit. Sci. Polit.* **50**, 456–460 (2017).
42. Penders, B. Ten simple rules for responsible referencing. *PLoS Comput. Biol.* **14**, e1006036 (2018).
43. Sumner, J. L. The gender balance assessment tool (GBAT): a web-based tool for estimating gender balance in syllabi and bibliographies. *PS Polit. Sci. Polit.* **51**, 396–400 (2018).
44. Lamont, J. & Favor, C. Distributive justice. in *The Stanford Encyclopedia of Philosophy*. Winter 2017 edn. <https://plato.stanford.edu/entries/justice-distributive/> (ed. Zalta, E. N.) (Stanford University, 2017).
45. Olsaretti, S. The idea of distributive justice. in *The Oxford Handbook of Distributive Justice* Vol. 1 (ed. Olsaretti, S.) <https://doi.org/10.1093/oxfordhb/9780199645121.013.38> (Oxford University Press, 2018).
46. Young, I. M. & Allen, D. S. *Justice and the Politics of Difference* (Princeton University Press, 2011).
47. Ahmed, S. *On Being Included: Racism and Diversity in Institutional Life* (Duke University Press, 2012).
48. Walker, M. U. *What is Reparative Justice? (The Aquinas Lecture 2010)* (Marquette University Press, 2010).
49. Anderson, E. *The Imperative of Integration* (Princeton University Press, 2010).
50. Gutiérrez, M. G., Niemann, Y. F., González, C. G. & Harris, A. P. (eds.) *Presumed Incompetent: The Intersections of Race and Class for Women in Academia* (University Press of Colorado, 2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Data collection. We drew data for this study from the WoS. This database indexes neuroscience journals according to the Science Citation Index Expanded, and we selected the neuroscience journals with the five highest Eigenfactor scores for this study. Eigenfactor scores give a count of incoming citations, where citations are weighted by the impact of the citing journal. Therefore, this measure roughly mimics the classic version of Google page rank and attempts to characterize the influence a journal has within its field³⁸. The journals selected were *Brain*, *Journal of Neuroscience*, *Nature Neuroscience*, *NeuroImage* and *Neuron*.

All articles published between 1995 and 2018 were downloaded, and articles classified as articles, review articles or proceedings papers that were labeled with a digital object identifier (DOI) were included in the analyses. The data downloaded for each paper from WoS included author names, reference lists, publication dates and DOIs, and we obtained information on each paper's referencing behavior by matching DOIs contained within a reference list to DOIs of papers included in the dataset.

Although authors' last names were included for all papers, authors' first names were only regularly included in the data for papers published after 2006. For all papers published in or before 2006, we searched for authors' first names using CrossRef API. When first names were not available on CrossRef, we searched for them on the journal webpage for the given article. To minimize the number of papers for which we only had access to authors' initials, to remove self-citations and to develop a co-authorship network, we implemented a name disambiguation algorithm.

Author name disambiguation. To minimize missing data and allow for name gender assignment and author matching across papers, we implemented an algorithm to disambiguate authors for whom different versions of their given name or initials were available across papers. We began by separating first and last names according to the method used by the given source (for example, WoS typically used 'last, first; last, first'). We then identified cases in which only initials were available following the previously described searching steps by marking authors for whom the first name entry contained only uppercase letters (as we found that many initials-only entries did not contain periods).

For each case, we collected all other entries that contained the same first or middle initials and the same last name. If only one unique first or middle name matched the initials of the given entry, or if distinct matches were all variants of the same name, we assigned that name to the initials. If multiple names in the dataset fit the initial and last name combination of the given entry, we did not assign a name to the initials. For example, if an entry listed an author as R. J. Dolan, and we found matches under Ray J. Dolan and Raymond J. Dolan, we replaced the R. J. Dolan entry with the more common completed variant. If, instead, we found matches under Ray J. Dolan and Rebecca J. Dolan, we did not assign a name to the original R. J. Dolan entry.

Next, we matched different name variants for the sake of tracking individual authors across their papers. To find and connect variants, we searched for instances of author entries with matching last names and either the same first name or first names that were listed as being commonly used nicknames according to the Secure Open Enterprise Master Patient Index⁵¹. If no matches fit that description, the name was retained. If one match occurred more commonly, the less common variant was changed to the more common variant. If multiple matches did not have any conflicting initials (some having a middle initial and others not having a middle initial was not considered conflicting), then less common variants were changed to the more common variant. If multiple matches had conflicting initials (for example, Ray Dolan being matched to both Raymond S. Dolan and Raymond J. Dolan), the target name was not changed.

There are three primary ways that incorrect author disambiguation could impact the results presented here. First, inability to link initials to an author's first name would yield missing data for papers that only included the author's initials. These papers would then not be included in the analyses as either cited or citing papers. Second, inability to link two versions of an author's name (for example, Ray Dolan and Raymond Dolan) would lead to the inclusion of some self-citations into the author's analyzed reference lists. This could lead to slightly inflated rates of authors citing other authors of the same gender, although sensitivity analyses suggest that this effect is essentially nonexistent in the present data (Supplementary Text and Supplementary Table 5). Third, incorrectly linking author A to author B would lead to the unnecessary removal of some citations (that is, any of author B's references to author A's work would be removed as self-citations). Although this is likely rare, its occurrence would lead to slightly decreased rates of authors citing other authors of the same gender.

Author gender determination. For authors with available first names, gender was assigned to first names using the 'gender' package in R⁵² with the Social Security Administration (SSA) baby name dataset. For names that were not included in the SSA dataset, gender was assigned using Gender API (<http://gender-api.com/>), a paid service that supports roughly 800,000 unique first names across 177 countries. We assigned 'man' ('woman') to each author if their name had a probability ≥ 0.70 of belonging to someone labeled as 'man' ('woman') according to a given source³⁵. In the SSA dataset, man and woman labels correspond to the sex assigned to

children at birth; in the Gender API dataset, man and woman labels correspond to a combination of sex assigned to children at birth and genders detected in social media profiles. Gender could be assigned to both the first and last author of 88% of the papers in the dataset. Of the 12% of papers with missing data, the first or last author's name either had uncertain gender (7%) or was not available (5%). We performed the following analyses using the articles for which gender could be assigned with high probability to both authors ($n = 54,225$), and sensitivity analyses conducted on the full data can be found in Supplementary Table 3.

To determine the extent of potential gender mislabeling, we conducted a manual study on a sample of 200 authors. The relative accuracy of the automated determination procedure at the level of both individual authors (accuracy ≈ 0.96 ; Supplementary Table 1) and article gender categories (accuracy ≈ 0.92 ; Supplementary Table 2) is presented in the Supplementary Information. Because errors in gender determination would break the links between citation behavior and author gender, any incorrect estimation in the present data likely biases the results towards the null.

Interpretation of author gender assignments. In gender theory, sex often refers to physical attributes, as determined anatomically and physiologically, while gender often refers to a self-identity, as expressed behaviorally and in a sociocultural context⁵³. In our analysis, the term 'gender' does not directly refer to the sex of the author, as assigned at birth or chosen later, nor does it directly refer to the gender of the author, as socially assigned or self-chosen. The term 'gender', in our analysis, is a function of the probability of assigned gendered names. By 'woman', we mean an author whose name has a probability ≥ 0.70 of being given to a child assigned female at birth or belonging to someone identifying as a woman on social media; likewise, by 'man', we mean an author whose name has a probability ≥ 0.70 of being given to a child assigned male at birth or belonging to someone identifying as a man on social media. The author's actual sex or gender is not identified.

Removal of self-citations. For this study, self-citations were removed from all analyses of gendered citation behavior. Although self-citations themselves have been found to have relevant gendered properties⁵⁴, their removal in this study allowed us to isolate more comparable external citation behaviors of men and women in the field. We further explored the role of self-citations in this data, specifically, the impact of including self-citations on the main results (Supplementary Fig. 1), and the relative prevalence of self-citations across author genders (Supplementary Table 6).

For the primary analyses, we defined self-citations as papers for which either the cited first or last author was the first or last author on the citing paper. While this definition is somewhat restrictive, it is the only type of self-citation for which the author gender of the cited paper is necessarily determined by the author gender of the citing paper. In addition, we demonstrated that using broader definitions of self-citation had little to no impact on the results, specifically, when the entire author list of the citing or cited paper was considered in the definition of self-citations (Supplementary Tables 7 and 8).

Statistical analysis. Many analyses conducted in this study relied on comparisons between observed citation behavior and the rates at which MM, WM, MW and WW papers would be expected to appear in reference lists if gender was irrelevant. To obtain expected rates that account for various characteristics that may be associated with gender, we fit a GAM on the multinomial outcome {MM, WM, MW and WW} in which the model's features were (1) the month and year of publication, (2) the combined number of publications by the first and last authors, (3) the number of total authors on the paper, (4) the journal in which it was published and (5) whether it was a review paper. When this model is then applied to each paper, it yields a set of probabilities that the paper belongs to the MM, WM, MW and WW categories. Importantly, this model does not predict the number of citations given to individual papers. Instead, it facilitates the calculation of the rates at which different gender categories would be expected to appear in reference lists if author gender were independent of citation rates, conditional on the other characteristics in the model. The GAM was fit using the mgcv package in R⁵⁵, using penalized thin-plate regression splines for estimating smooth terms of publication date, author experience and team size. For the primary analyses, univariate thin-plate splines were used for the smooth terms, and no interactions between variables were included in the model. Supplementary Fig. 2 shows the effect of fitting a more complex model on the main results. The results demonstrate that the incorporation of interaction terms and multivariate splines has little impact on the outcomes of interest.

Estimates in this study are presented with a CI, a *P* value or both. CIs in this study were calculated by bootstrapping citing papers (that is, randomly sampling citing papers with replacement). As opposed to bootstrapping individual instances of citations, this method maintained the dependence structure of the clusters of cited articles within citing articles. The null model used to obtain *P* values was derived from the randomization of author gender categories of cited papers. Randomization was carried out by probabilistically drawing new gender categories for each paper according to their GAM-estimated gender probabilities. Randomly sampling gender categories for each paper, therefore, produces a null model in which cited author gender is conditionally independent of citation rates and citing

author gender (conditional on the characteristics included in the GAM model), while the structure of the citation graph and the long-tailed nature of the citation distribution are both preserved. A total of 10,000 randomizations were carried out to calculate P values (see Supplementary Table 10 for summary measures of estimates' null distributions across randomizations and Supplementary Fig. 3 for a visualization of the null distributions). Because significance was assessed for multiple primary comparisons, all presented P values were corrected according to the Holm–Bonferroni method⁵⁶.

In the following sections, we describe the formal statistical analysis that we used to address the four distinct hypotheses. In each subsection, we state the hypothesis first, followed by the analysis used to test it. All hypotheses were tested for the set of articles published between 2009 and 2018. We decided to specifically consider reference lists from the past 10 years to ensure that estimates of over/undercitation were reflective of current behavior, were not a result of aggregating over disparate eras of neuroscience research and contained enough previous citable papers to represent meaningful and stable measures of behavior.

Hypothesis 1: *the overall citation rate of women-led papers will be lower than expected given relevant characteristics of papers.* To test this hypothesis, we first estimated the expected number of citations given to each author gender category. We calculated this expectation by summing over the GAM-estimated probabilities for all papers contained within the reference lists of citing papers. These totals, therefore, reflect the expected number of citations given to MM, WM, MW and WW papers if author gender were conditionally independent of citation behavior, given the paper characteristics included in the model described above.

To calculate the observed number of citations given to each group, we simply summed over the {MM, WM, MW and WW} dummy variable for all of the papers contained within the reference lists of papers published between 2009 and 2018. These values were compared by calculating the percentage difference from expectation for each author gender group. For example, for WW papers, this percentage change in citation would be defined as:

$$\Delta_{WW} = \frac{\text{obs}_{WW} - \text{exp}_{WW}}{\text{exp}_{WW}}$$

where obs_{WW} is the number of citations given to WW papers between 2009 and 2018, and exp_{WW} is the expected number of citations given to WW papers between 2009 and 2018.

Notably, performing the summation over all citations resulted in the upweighting of articles with many citations and the downweighting of articles with few citations. This approach helps to improve the stability of the estimates but could potentially be sensitive to high-influence observations. To determine the impact of this decision, we conducted sensitivity analyses in which we used the mean of article-level over/undercitation and found little difference between the two estimation strategies (Supplementary Text and Supplementary Table 9).

Hypothesis 2: *undercitation of women-led papers will occur to a greater extent within men-led reference lists.* To test this hypothesis, we used very similar metrics to those described above. The primary difference is that, instead of calculating the observed and expected citations by summing over the citations within all reference lists between 2009 and 2018, we performed those summations separately for reference lists in papers with men as first and last author (MM papers) and papers with women as first or last author (WUW papers). For example, the over/undercitation of WW papers within the reference lists of MM papers were defined as:

$$\Delta_{WW}^{(MM)} = \frac{\text{obs}_{WW}^{(MM)} - \text{exp}_{WW}^{(MM)}}{\text{exp}_{WW}^{(MM)}}$$

where $\text{obs}_{WW}^{(MM)}$ is the total number of citations given to WW papers within MM reference lists, and $\text{exp}_{WW}^{(MM)}$ is the expected number of citations given to WW papers within MM reference lists.

Hypothesis 3: *undercitation of women-led papers will decrease over time but at a slower rate within men-led reference lists.* As there are four separate measures representing over/undercitation of each author group, we calculated the change in the overcitation of men over time using a simple measure of the absolute difference between the observed and expected proportion of MM papers cited. This measure of change is given by:

$$\delta_{MM,\text{year}} = \frac{\text{obs}_{MM,\text{year}} - \text{exp}_{MM,\text{year}}}{\text{obs}_{\text{year}}}$$

where obs_{year} is the total number of citations within a given year, $\text{obs}_{MM,\text{year}}$ is the number of citations given to MM papers in a specific year and $\text{exp}_{MM,\text{year}}$ is the expected number of citations given to MM papers in a specific year. The change in the overcitation of men over time is estimated using a linear regression of $\delta_{MM,\text{year}}$ on year, the CI of this estimate was obtained using the article bootstrap procedure, and significance was assessed using the graph-preserving null model.

Similarly, to estimate the change in overcitation of MM papers separately within MM and WUW reference lists, we defined group-specific measures of yearly overcitation. For example, overcitation of MM papers within MM reference lists for a specific year would be given by:

$$\delta_{MM,\text{year}}^{(MM)} = \frac{\text{obs}_{MM,\text{year}}^{(MM)} - \text{exp}_{MM,\text{year}}^{(MM)}}{\text{obs}_{\text{year}}^{(MM)}}$$

where $\text{obs}_{\text{year}}^{(MM)}$ is the total number of citations within MM reference lists in a specific year, $\text{obs}_{MM,\text{year}}^{(MM)}$ is the number of citations given to MM papers within MM reference lists in a specific year, and $\text{exp}_{MM,\text{year}}^{(MM)}$ is the expected number of citations given to MM papers within MM reference lists in a specific year.

Hypothesis 4: *differences in undercitation between men-led and women-led reference lists will be partly explained by the structure of authors' social networks.* To test this hypothesis, we developed a temporal co-authorship network in which nodes were individual authors (only authors who appeared as first or last author in at least one paper in the dataset were included), and binary edges represented the fact that two authors had appeared on at least one paper together before a given date. Of interest next was to estimate the relationship between authors' local network composition and their citation behavior. Because citation behavior occurs at the level of a reference list within a specific paper with both a first and a last author (rather than at the level of a single node or author), we sought to define two measures of local network composition at the paper level. For the purposes of these analyses, we considered a paper to be the set $\{a_p, a_l, m\}$, where a_p is the first author, a_l is the last author and m is the month of publication. We then defined a paper's local neighborhood of authors, N_p^p , to be the authors that are connected by shared publication to either a_p or a_l before month m . We also define a paper's local neighborhood of papers, N_p^p , to be the union of all papers authored by anyone within N_p^p before month m .

The two measures of local network composition are man-author over-representation and MM-paper over-representation. We defined man-author over-representation as the difference between the proportion of men within a paper's local author neighborhood, N_p^p , and that of the overall network. For paper p , this measure is thus given by:

$$\text{MA}_{\text{or}}(p) = \pi_{M,N_p^p} - \pi_M$$

where π_M is the proportion of men in the full co-authorship network, and π_{M,N_p^p} is the proportion of men within paper p 's local author neighborhood. Similarly, we defined MM-paper over-representation as the difference between the proportion of MM articles within a paper's local paper neighborhood, N_p^p , and that of the overall network. For paper p , this measure is given by:

$$\text{MMP}_{\text{or}}(p) = \pi_{MM,N_p^p} - \pi_{MM}$$

where π_{MM} is the overall proportion of MM articles within the data, and π_{MM,N_p^p} is the proportion of MM articles within paper p 's local paper neighborhood.

To estimate the relationship between these metrics and the degree of overcitation of men within reference lists, we defined a paper-level measure of the absolute difference between the observed and expected proportion of MM papers. Similarly to the previously described $\delta_{MM,\text{year}}^{(MM)}$ measure that quantified the overcitation of MM papers within all MM reference lists from a given year, here we defined a measure of overcitation within an individual paper p . It is given by:

$$\delta_{MM}^{(p)} = \frac{\text{obs}_{MM}^{(p)} - \text{exp}_{MM}^{(p)}}{\text{obs}^{(p)}}$$

where $\text{obs}_{MM}^{(p)}$ is the number of MM citations within paper p 's reference list, $\text{exp}_{MM}^{(p)}$ is the expected number of MM citations within paper p 's reference list based on the GAM-estimated assignment probabilities of each cited paper and $\text{obs}^{(p)}$ is the total number of candidate citations within paper p 's reference list.

The relationships between $\delta_{MM}^{(p)}$, $\text{MMP}_{\text{or}}(p)$, $\text{MA}_{\text{or}}(p)$ and {MM, WM, MW and WW} are estimated using weighted quantile regression, with the MM overcitation metric, $\delta_{MM}^{(p)}$, as the outcome. We performed quantile regression because of the bounded and skewed nature of the $\delta_{MM}^{(p)}$ measure, and the results of a sensitivity analysis using linear regression can be found in Supplementary Table 11. We defined the weights to be equal to the number of candidate citations within a given paper's reference list; this choice gives higher weight to papers for which the outcome is more stable. Results from an unweighted model can be found in Supplementary Table 9. We also selected the τ value of the quantile regression formula to be 0.5, resulting in a model fit to the median of the outcomes. CIs were again obtained by the article bootstrap method, and significance was assessed using the graph-preserving null model.

Efforts to expand the methodological scope of this study could build on the above analysis of collaboration networks. While the present study investigates the role of network structure in citation behavior, previous work has found that the gender^{57,58}, ethnic⁵⁹ and international^{60,61} composition of the collaboration networks of scholars can be related to scientific impact and career success. Such effects in

these data could provide additional information about mechanisms and potential points of intervention. Additionally, future work could incorporate longitudinal within-author analyses of co-authorship networks and citation behavior. Such analyses may provide a better understanding of how author practices change over time and are impacted by the gender of their co-authors; this knowledge could facilitate more individualized recommendations.

Citation gender diversity statement. The gender balance of papers cited within this work was quantified using a combination of automated gender-api.com estimation and manual gender determination from authors' publicly available pronouns. Among the 60 cited works that had named authors, 28% ($n=17$) were MM, 12% ($n=7$) were WM, 13% ($n=8$) were MW and 47% ($n=28$) were WW.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data and materials for this study have been deposited in an Open Science Framework repository and can be accessed at <https://osf.io/h79g8/>.

Code availability

Code for reproducing presented estimates and figures can be accessed at <https://osf.io/h79g8/>, and code that demonstrates the full sampling, processing and analysis pipeline can be accessed at <https://github.com/jdvor/gendercitation>.

References

51. Toth, C., Durham, E., Kantarcioğlu, M., Xue, Y. & Malin, B. SOEMPI: a secure open enterprise master patient index software toolkit for private record linkage. *AMIA Annu. Symp. Proc.* **2014**, 1105–1114 (2014).
52. Blevins, C. & Mullen, L. Jane, John ... Leslie? A historical method for algorithmic gender prediction. *Digit. Humanit. Q.* **9**, 2015.
53. Fausto-Sterling, A. *Sexing the Body: Gender Politics and the Construction of Sexuality* 1st edn, (Basic Books, 2000).
54. 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**, 237802311773890 (2017).
55. Wood, S. N. *Generalized Additive Models: An Introduction with R* 2nd edn, (Chapman and Hall/CRC, 2017).
56. Sture, H. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6**, 65–70 (1979).
57. Jadidi, M., Karimi, F., Lietz, H. & Wagner, C. Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. *Adv. Complex Syst.* **21**, 1750011 (2018).
58. Yang, Y., Chawla, N. V. & Uzzi, B. A network's gender composition and communication pattern predict women's leadership success. *Proc. Natl Acad. Sci. USA* **116**, 2033–2038 (2019).
59. AlShebli, B. K., Rahwan, T. & Woon, W. L. The preeminence of ethnic diversity in scientific collaboration. *Nat. Commun.* **9**, 5163 (2018).
60. Uhly, K. M., Visser, L. M. & Zippel, K. S. Gendered patterns in international research collaborations in academia. *Stud. High. Educ.* **42**, 1–23 <https://doi.org/10.1080/03075079.2015.1072151> (2015).
61. Zippel, K. S. *Women in Global Science: Advancing Academic Careers Through International Collaboration* (Stanford University Press, 2017).

Acknowledgements

We thank D. Lydon-Staley and D. Zhou for constructive comments on an earlier version of this manuscript. R.T.S. would like to acknowledge support from the National Institute of Neurological Disorders and Stroke (R01 NS085211 and R01 NS060910). D.S.B. acknowledges support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation and an NSF CAREER award (PHY-1554488). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Author contributions

Conceptualization: J.D.D., K.A.L., R.T.S. and D.S.B.; methodology: J.D.D., K.A.L., E.G.T., R.T.S. and D.S.B.; data curation: J.D.D.; formal analysis: J.D.D.; writing: J.D.D., P.Z. and D.S.B. (original draft) and J.D.D., K.A.L., E.G.T., P.Z., R.T.S. and D.S.B. (review and editing); funding acquisition: R.T.S. and D.S.B.; supervision: R.T.S. and D.S.B.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41593-020-0658-y>.

Correspondence and requests for materials should be addressed to D.S.B.

Peer review information *Nature Neuroscience* thanks Clarissa Bauer-Staub, Katherine Button, Catherine Hobbs, Russell Poldrack and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data for this study were collected from the Web of Science (webofknowledge.com).

Data analysis

R version 3.6.0 was used for the analyses presented in this study. The packages used for analysis are as follows: bibliometrix, rvest, dplyr, parallel, pbmcapply, rjsonio, textclean, stringr, tidyverse, rjson, mgcv, ggplot2, patchwork, boot, data.table, and quantreg.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data that support the findings of this study have been deposited in an Open Science Framework repository and can be accessed at <https://osf.io/h79g8/>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|-------------------|--|
| Study description | This study is a quantitative and cross-sectional assessment of citation patterns in top neuroscience journals. Within the sample of articles (see "Research sample" below), we obtain probabilistic estimates of authors' gender, find connections between citing and cited papers, locate and remove self-citations, and study the links between authors' genders and their role as objects/agents of undercitation. |
| Research sample | The sample is research articles and review articles published between 1995 and 2018 within one of five neuroscience journals (Nature Neuroscience, Journal of Neuroscience, Neuron, NeuroImage, and Brain). These journals were pre-selected, as they represent the five journals with the highest Eigenvalue scores in the neuroscience category on Web of Science. Research articles and review articles were selected as they were seen by the authors as being the two primary outlets through which previous research is cited, discussed, and built upon, and their labels have consistent meanings across journals. |
| Sampling strategy | Data for all articles meeting the above inclusion criteria were directly obtained through Web of Science. Therefore, no sampling strategy was required. |
| Data collection | Data were collected through Web of Science queries of publicly available bibliometric information. |
| Timing | Data were collected between February 3, 2019 and February 11, 2019. |
| Data exclusions | 7190 of 61416 articles (12%) were excluded due to pre-determined exclusion criteria. Namely a lack of ability to confidently adjudicate author gender. Sensitivity analysis were conducted on the full sample (including missing data) and are available in the Supplementary Information. |
| Non-participation | N/A. There were no human participants in this study. |
| Randomization | N/A. There was no randomization of participants in this study. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | | Methods | |
|-------------------------------------|--|-------------------------------------|---|
| n/a | Involved in the study | n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies | <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines | <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology | <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms | | |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants | | |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data | | |