nature physics

Perspective

https://doi.org/10.1038/s41567-022-01770-1

Citation inequity and gendered citation practices in contemporary physics

In the format provided by the authors and unedited

Supplementary Information: Citation inequity and gendered citation practices in contemporary physics

Erin G. Teich,[1] Jason Z. Kim,[1] Christopher W. Lynn,[2,3] Samantha C. Simon,[4] Andrei A. Klishin,[1] Karol P. Szymula,[5] Pragya Srivastava,[1] Lee C. Bassett,[6] Perry Zurn,[7] Jordan D. Dworkin,[8,9] and Dani S. Bassett[4,1,6,10,11,12,13]

[1)] *Department of Bioengineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104 USA*
[2)] *Initiative for the Theoretical Sciences, Graduate Center, City University of New York, New York, NY 10016, USA*
[3)] *Joseph Henry Laboratories of Physics and Lewis–Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA*
[4)] *Department of Physics & Astronomy, College of Arts & Sciences, University of Pennsylvania, Philadelphia, PA 19104 USA*
[5)] *School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642 USA*
[6)] *Department of Electrical & Systems Engineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104 USA*
[7)] *Department of Philosophy & Religion, American University, Washington, D.C. 20016 USA*
[8)] *Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY 10019 USA*
[9)] *New York State Psychiatric Institute, New York, NY 10032 USA*
[10)] *Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 USA*
[11)] *Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 USA*
[12)] *Santa Fe Institute, Santa Fe, NM 87501 USA*
[13)] *To whom correspondence should be addressed: dsb@seas.upenn.edu*

(Dated: 9 August 2022)

S1. SUPPLEMENTARY METHODS: DATA ACQUISITION

A. Journal selection

In order to broadly characterize citation behavior within and across the subdisciplines of physics, we selected a list of peer-reviewed journals guided by several criteria. First, the journals needed to cover all major subfields of physics, as defined by the breakdown of the *Physical Review* family of journals and the *Web of Science* (WoS) database categories. Second, within each subfield we selected the central journals based on their Eigenfactor (EF) score[1] (a measure that rates journals according to incoming citations, weighted by the impact of those citations' journals) as reported by InCites Journal Citation Reports for the year 2018. Third, we aimed to represent each subfield by an equal number of journals, while allowing the number of papers to differ. To balance these three considerations, we accessed lists of journals on WoS according to all pre-defined "journal categories" that we judged to correspond to each physics subfield as defined by the *Physical Review* family of journals. We also included an Applied Physics/Nanoscience subfield, which did not correspond to a *Physical Review* journal. We ranked each list of journals according to EF score, from highest to lowest. We then took the top 5 journals according to EF score from each list, after some journal exclusions. If a journal appeared in multiple lists, we excluded it from all redundant lists except its best content-based fit. We also excluded journals whose content was not physics-centric enough, or whose content was inappropriate for the relevant physics subfield. For example, the WoS journal category *Physics, Atomic, Molecular & Chemical* contained several high-EF chemistry journals, which we excluded when creating our atomic, molecular, and optical physics list of journals. Similarly, the WoS journal category *Biophysics* contained several high-EF biology and biochemistry journals, which we excluded when creating our soft matter physics list of journals. And the WoS journal category *Physics, Fluids & Plasmas* contained several high-EF fluid mechanics and exclusively plasma physics journals, which we excluded when creating our nuclear physics list of journals. We also found that our procedure resulted in overlapping lists of journals for the condensed matter and applied physics/nanoscience subfields, so we disambiguated these lists according to the stated scope of each journal.

This initial search resulted in 7 non-overlapping lists of 5 journals each, for a total of 35 journals. These journals are listed in Fig. 1b, and the number of articles contained in our data set in each journal per year is shown in Table S1. We emphasize that our choice of journals with high Eigenfactor score was deliberate, as we purposefully sought in this work to measure citation imbalance in journals considered "impactful" within their communities. These journals

presumably contain the most widely-read and discussed articles within each community, and these articles' imbalances can be considered the most visible and influential within each community.

B. Data collection

We downloaded all papers published in the 35 chosen journals between 1995 and 2020 from the *Web of Science* database. We then selected all papers classified as original research articles or review articles, and extracted information on author names, reference lists, publication dates, and DOIs. Each paper's citation behavior was obtained by matching DOIs contained within its reference list to DOIs of papers in our dataset. Authors' last names were included for all papers; however, for a portion of papers, authors' first names were not included in the database. First names are necessary for our name-based author gender categorization scheme (see Section S2 C), so we attempted to find these missing first names in two ways: (i) Searching for author first names using Crossref's API, and (ii) Implementing a name disambiguation scheme (see Section S2 A) whereby authors' initials and/or other name abbreviations are matched to and replaced by their full first names if possible.

S2. SUPPLEMENTARY METHODS: DATA PREPROCESSING

A. Author name disambiguation

To maximize the number of analyzable papers for subsequent author gender categorization by first names, we employed a method of disambiguating authors for whom several versions of their name or initials were available across papers, followed by assignment of the most complete version of each author's name to all papers they authored. Due to the size of our dataset, we performed our name disambiguation procedure in parallel on isolated subsets of papers, grouped according to their initially-defined subfields shown in Fig S1. This method involved several steps. First, we identified all cases for which first and/or last authors' first names consisted only of initials by isolating all authors' first names for each paper, and then flagging the first and last of those which contained only uppercase letters. Next, for each case of a flagged first name that only contained initials, we gathered all other name instances with the same first/middle initials and the same last name. If only one unique first/middle full name matched the initials-only entry, or if all distinct full name matches were variants of the same name, we assigned that name to the initials. However, if multiple unique first/middle full names matched the initials-only entry, we did not assign a name to the initials. For example, if an entry listed an author as A. A. Griffin, and we found matches under Abby A. Griffin and Abigail A. Griffin, we would replace the A. A. Griffin entry with Abigail A. Griffin. If instead we found matches under Abby A. Griffin and Arlene A. Griffin, we would not assign a name to A. A. Griffin. Initially, our dataset contained 463,538 initials-only first and/or last author name entries. Through the steps outlined above, we were able to assign full first names to 88,780 of these entries, for a success rate of $88,780/463,538 \approx 19.15\%$.

We employed a similar strategy to match name variants in order to more accurately determine authorship histories of individual authors. For every author entry, we first identified sets of corresponding author entries with matching last names and either the same first name or first names determined to be common nicknames of each other (e.g., Abby and Abigail) according to the Secure Open Enterprise Master Patient Index[2]. If no corresponding matches existed, we retained the author entry name. If corresponding matches did exist and one match occurred more often, the less common name variant was changed to the more common name variant. Similarly, if the corresponding matches did not have any conflicting initials, the less common initial variants were changed to the more common initial variants. A common non-conflicting scenario in this case was that some matches had middle initials and others had no middle initials. If multiple corresponding matches did have conflicting initials, however, the author entry name was not changed.

The name disambiguation scheme just described was only performed within paper subsets corresponding to subfields, in part because of the size of our data set. We emphasize, however, that overly aggressive name disambiguation across subfields is not ideal, and might actually introduce spurious name variant matches that affect the accuracy of our results. As an example, if an author named Sarah Ruth published papers in the nuclear physics subfield in the early 2000s, and another author named Sarah R. Ruth published papers in the soft matter subfield in 2018, an aggressive name disambiguation scheme applied across subfields might assign the middle initial R. to the nuclear physicist named Sarah Ruth. However, for common names, given the size and scope of our data set and the massively distinct training and expertise needed to work in different subfields of physics, it is not a given that these are the same person. Our less aggressive name disambiguation scheme applied only within subfields is thus a way, albeit limited, to avoid these false positive cases.

We also note that, although name disambiguation was only performed within subfields, self-citations (defined in Section S2E) of each paper were subsequently identified as citations whose author names were identical to citing author names, regardless of paper subfield. For example, if an author team featuring Sarah Ruth cited any other paper written by an author team featuring Sarah Ruth, it was identified as a self-citation regardless of either paper's subfield.

B. Estimation of publication month

While the year of publication was available for all 1,067,276 papers, the publication month was not available for 38,423 papers ($\approx 3.6\%$). In order to approximate the unknown month m_i for each paper i published in year y_i, we considered its lower and upper bounds. The lower bound was set to be the month of publication of the most recent paper cited by i if the most recent paper was published in year y_i, or January otherwise. The upper bound was set to be the month of publication of the first paper to cite i if the first paper was published in year y_i, or December otherwise. We then approximated m_i as the midpoint between the upper and lower bounds. To assess the validity of this approach, and to understand the associated uncertainty, we performed the same analysis on the $1,067,276 - 38,423 = 1,028,853$ papers for which the publication month was available. In this test case, we found that the average absolute error between the true month and the estimated month was ≈ 2.27 months. In contrast, the average absolute error of naively guessing a publication month between June and July was ≈ 2.98, indicating that our method provides a reasonable approximation.

C. Name-based assignment of author gender categories

We assigned "author gender categories" to papers according to first names of papers' first and last authors. Gender was assigned to authors with available first names using Gender API, a paid service that includes statistics for approximately 6 million unique first names across 191 countries at the time of writing[3]. We assigned the label 'man' ('woman') to each author if their first name had a probability ≥ 0.7 of belonging to someone labeled 'man' ('woman') according to our sources[4]. Labels are assigned in the Gender API dataset according to a combination of sex assigned to children at birth or chosen by adults later, and gender detected in social media profiles.

Our dataset includes $n = 1,067,276$ papers, for a total of n first authors and n last authors. Among them, we were able to assign gender labels to both first and last authors in 60% of papers, to only one of either first or last author in 27% of papers, and to neither the first nor last author in 13% of papers. Among all first and last authors with unassigned gender labels, 65.4% were due to publishing using initials. To ensure that omitting these authors in subsequent analyses would not significantly skew our results, we sought to estimate the gender label distribution among these authors. As a proxy for this author set, we examined the set of authors for which we uniquely matched initials to first names using our name disambiguation scheme, and successfully assigned author gender labels (see Section S2 A). Of these 79,592 authors, 66,255 ($\approx 83.24\%$) were assigned the label 'man' by our algorithm and 13,337 ($\approx 16.76\%$) were assigned the label 'woman' (see the *Supplementary Information* for details). These numbers are consistent with the ratios of assigned gender labels of all first and last authors in our dataset, of which $\approx 87.92\%$ were labeled 'man' and 12.07% were labeled 'woman.' Hence, as far as we can measure, the distribution of genders in unassigned author names is not radically different than that of assigned author names.

We then subdivided papers into author gender categories according to assigned gender labels of first and last authors. If the first and/or last author's name was assigned the label 'woman,' we categorized the paper as W||W. To increase statistical power, we included all papers in this category with at least one woman-assigned first/last author, even if the other author could not be assigned a gender label according to our methods. If the first and last author's name was assigned the label 'man,' we categorized the paper as MM.

We emphasize that, although the phrase "author gender category" contains the word "gender," it need not capture the true gender identities of all authors. Instead, it expresses a statistical correlation: By 'woman' we mean an author whose name has a probability greater than or equal to 0.70 of belonging to someone identifying as a woman on social media or in federal documents; likewise, by 'man,' we mean an author whose name has a probability greater than or equal to 0.70 of belonging to someone identifying as a man on social media or in federal documents. True gender identity could only be learned through careful manual research of self-attested gender identity, or already known through kinship or conversation, and is not accessible via an automated analysis pipeline like the one used in this paper. However, the author gender category is nevertheless a notably useful proxy of gender for the purposes of this paper, because it expresses a correlation between name and gender. Names greatly influence perceptions of gender identity[5], with well-known implications for a person's perceived merit as a scientist[6]. These perceptions have marked power to shape citation behavior, irrespective of authors' true gender identities.

D. Authors of unassigned gender

Through the process of determining author gender categories, we were unable to assign the genders of many authors. In our final dataset, among the 1,067,276 papers, we were unable to assign the genders of 313,535 first authors, and 259,636 last authors. Among them, we were unable to assign the genders of both the first and last authors for 142,601 papers. This final dataset includes full names for authors whose initials were uniquely matched to full names in the dataset.

To understand the cause behind these unassigned genders, we quantify how many of these unassigned author names originate from authors who publish papers under initials. Among the 313,535 first authors with unassigned genders, 204,578 are unassigned due to publishing under an initialed first name. Among the 259,636 last authors with unassigned genders, 170,180 are unassigned due to publishing under an initialed first name.

To estimate the gender statistics among authors who publish under initials, we carried out a unique matching of initialed names to known full names. First, we made a database of non-initialed first names and last names $\mathcal{A}$ across all authors and papers in our dataset. Next, for each first or last author with an initialed first name, we used regular expressions to match their initials to $\mathcal{A}$. If the initials uniquely matched one first and last name pair, then the initials were replaced with the non-initialed first and last name. Through this method, we uniquely identified the first and last names of 88,780 first and last authors, among which we assigned genders for 79,592 authors. Among these authors with uniquely identified initials and assigned genders, 66,255/79,592 ($\approx 83.24\%$) were assigned "man," and 13,337 ($\approx 16.76\%$) were assigned "woman." These numbers are consistent with the authors of assigned genders in the entire dataset. Specifically, among the 1,067,276 - 313,535 = 753,741 first authors with assigned genders and 1,067,276 - 259,636 = 807,640 last authors with assigned genders (for a joint total of 1,561,381 first and last authors with assigned genders), we assigned 1,372,849 / 1,561,381 ($\approx 87.92\%$) "man," and 188,532 / 1,561,381 ($\approx 12.07\%$) "woman."

E. Reference list cleaning procedure

We pared down each paper's reference list into a "clean" version suitable for our analysis, which we used for all investigations of citation behavior (unless indicated otherwise). We determined citations according to DOI: In each paper's reference list, we gathered the set of cited DOIs and determined which of those matched any other paper's DOI in our data set. Of the 1,067,276 papers in our data set, 15,207 papers had missing DOIs, or $\approx 1.42\%$ of all papers.

From each reference list, we removed self-citations, citations to papers not in our dataset, and citations to papers with authors whose names could not be assigned to a gender category. We chose to remove self-citations from all analyses of citation behavior in order to focus on the influence of perception of other authors' gender category on external citation behavior. Self-citations were defined somewhat restrictively as references to papers for which either the cited first or last author was also the citing first or last author. This choice reflected the overall focus of the paper on the citation statistics of citing and cited teams consisting of first and last authors on all papers. We note that in a related study of citation behavior in neuroscience journals, it was shown that including self-citations did not result in meaningful differences in overall over-/under-citation trends[7]. Additionally, alternate (broader) definitions of self-citations were shown to yield highly similar results to our more restrictive definition of papers for which either the cited first or last author was also the citing first or last author[7]. Section S4H explores the influence of including self-citations in the analyses presented in this paper.

S3. SUPPLEMENTARY METHODS: ANALYSIS

A. Probability estimation of author gender category according to paper characteristics

To quantify over-/under-citation behavior, we first developed a gender-blind null model predicting the probability that papers were written by MM or W||W author teams according to a set of paper characteristics enumerated below. Then, over any set of reference lists, we could tally the number of cited papers written by MM or W||W author teams and compare these quantities to the expected numbers given by the gender-blind null model. The characteristics we used to predict MM or W||W authorship for each paper were (i) month and year of publication, (ii) first and last authors' combined number of papers in the dataset, (iii) total number of authors, (iv) publishing journal, and (v) categorization as a research or review article (Fig. 2a). Specifically, we fit a generalized additive model (GAM) on the binomial outcome {MM, W||W} with predictive features as defined above. We fit this model to all papers in our dataset whose authors' names could be assigned gender categories. To fit the GAM, we utilized the 'mgcv' package in R[8], using penalized thin plate regression splines for estimating smooth terms of features (i), (ii), and (iii) described

above. We note that we used the logarithm of feature (iii) and a winsorized version of feature (ii), capped at 300 (representing the top 0.6% of papers), to ensure that we fit the GAM successfully. Univariate thin plate splines were used for the smooth terms, and no interactions between variables were included in the model.

For each article, the GAM then yields a predicted probability of MM or W||W authorship; we interpret and utilize these probabilities as approximations of the proportion of similar articles (i.e., articles with comparable values of the above characteristics) written by each group. We then aggregate these probabilities across reference lists, allowing us to calculate the proportions of MM and W||W citations that would be expected if references were drawn in a gender-agnostic manner from pools of characteristic-matched papers. As a result, we can assess citation imbalances that account for the potentially confounding factors represented by those characteristics. For example, suppose an author team cites more papers from *Physical Review Letters* (PRL) than any other journal in the data set. Simply calculating the fraction of that author team's cited papers that are MM, and comparing that fraction with the overall field's MM paper proportion, would represent an estimate of citation imbalance confounded by that author team's preference to cite PRL. Since PRL skews toward a higher proportion of MM papers, the author team's higher proportion of cited MM papers could potentially be explained by the simple fact that they cite in a gender-agnostic manner from PRL. To address this issue, we use the GAM to produce estimates for the PRL-specific (and other paper characteristic-specific) MM and W||W paper proportions, so that we can compare author teams' cited MM (and W||W) proportions to more specific expected proportions in order to compute less confounded estimates of citation imbalance.

Once fit, the GAM predicts author gender category according to paper characteristics in a manner that is consistent with observed author gender categories. Figs. S1a and b show distributions of the GAM-predicted probabilities that papers are written by MM and W||W teams, for all MM and W||W papers in the data set. GAM-predicted probabilities that papers are written by MM teams are higher for papers actually written by MM teams (median 0.77) than for papers actually written by W||W teams (median 0.70). GAM-predicted probabilities that papers are written by W||W teams are higher for papers actually written by W||W teams (median 0.30) than for papers actually written by MM teams (median 0.23). Wilcoxon rank sum tests (or Mann-Whitney U tests) indicate that both GAM-predicted probability distributions for papers actually written by MM teams are shifted with respect to their corresponding GAM-predicted probability distributions for papers actually written by W||W teams ($p <$ 2.2e-16 in both cases).

The fitted GAM also shows variation in author gender category as a function of paper characteristics that is similar to observed variation. The left panels of Fig. S2 show partial effect plots for all paper characteristics included in the GAM. In other words, they show the relationship between each paper characteristic and the probability (specifically, the log odds ratio) that papers are written by W||W teams with respect to MM teams. The right panels of Fig. S2 show the actual fraction of papers written by MM and W||W teams as a function of the paper characteristics. For all paper characteristics, the partial effects modeled by the GAM generally track the actual trends observed in the data set. Predictions indicate that papers are more likely to be written by women over time, as the publication month from the earliest publication date in the data set ("month from base") increases (Fig. S2a). Papers are predicted most likely to be written by women in the journals *Astronomy & Astrophysics*, *Monthly Notices of the Royal Astronomical Society*, and *Soft Matter*, and least likely to be written by women in the journals *Nuclear Fusion*, *Reviews of Modern Physics*, and *Nature Photonics* (Fig. S2b). Papers are predicted to generally be less likely to be written by women as the first and last authors' combined number of papers in the data set, or author seniority, increases (Fig. S2c). Papers are predicted to generally be more likely to be written by women as the number of direct co-authors, or the log of paper team size, increases (Fig. S2d). Note that there is a dip in the partial effect modeled by the GAM at extremely high values of paper team size; this is an artifact of the spline estimation of the GAM, and occurs at values of team size that lie at the upper limit of the data range. And finally, review papers are predicted to be more likely written by women with respect to non-review papers (Fig. S2e). The proportion of null deviance in author gender category that is explained by paper characteristics is 6.1%. This suggests, perhaps unsurprisingly, that author gender cannot be reliably predicted from these five paper characteristics. Importantly, however, our goal is only to remove the potential confounding effects of these characteristics, not to create a gender prediction model. For example, if more men publish in Journal A, and Journal A is more highly cited regardless of gender, this could induce a spurious gender gap unless journal is accounted for when quantifying citation bias. By using this model, we are able to estimate and account for the relationships between author gender and paper characteristics, however large or small those relationships may be. Table S2 shows coefficients for linear terms in the GAM, errors of those terms, and p-values for the null hypothesis that each linear term is zero. Table S3 shows the effective degrees of freedom of all smoothed terms in the GAM, and p-values for the null hypothesis that each smooth term is zero. p-values are marked with $***$ if they are less than 0.001, $**$ if they are less than 0.01, and $*$ if they are less than 0.05. Details regarding all of these parameters and accompanying tests of statistical significance can be found in the documentation of the 'mgcv' package in R[8].

B. Calculation of over-/under-citation

For a given group of citing papers, we defined its over/under-citation of each author gender category as the percent difference of observed citations from gender-blind expectation. Over-/under-citation of MM papers, for example, is given by $(o_{MM} - e_{MM})/e_{MM} * 100$, where o_{MM} is the (observed) number of citations given to MM papers by the citing papers, and e_{MM} is the expected number of MM citations predicted by the gender-blind model described in the previous section. More specifically, e_{MM} is computed by summing over the GAM-estimated probabilities that each citation given by the group belongs to the author gender category MM.

C. Subfield delineation

To understand how citation practices vary between disciplines, we grouped journals into defined "subfields." The boundaries between subfields were drawn according to a combination of (i) pre-defined journal categories culled from the breakdown of the *Physical Review* family of journals and journal categories defined by *Web of Science*, and (ii) *post-hoc* citation network clustering to verify that journals within subfields cited each other to at least some degree. We also sought to maintain somewhat equitable numbers of journals within each subfield for subsequent analyses of citation behavior on the subfield level. The resultant subfields and their constituent journals are shown in Fig. 1b.

Initial journal selection according to consideration (i) above is described in Section S1 A. For consideration (ii) above, citation network visualization and clustering, we first built a directed citation network between the 35 journals in our dataset (Fig. S3, shown with the original subfield boundaries built according to consideration (i)). Each element J_{ij} of the matrix in Fig. S3 represents a directed citation flow (weighted edge) between citer journal (node) i and cited journal (node) j: $J_{ij} \equiv N_{ij}/\sum_j N_{ij}$, where N_{ij} is the number of citations given by p_i, the set of counted papers published in journal i, to p_j, the set of counted papers published in journal j. To most accurately capture the citation dynamics we analyzed in the main text, we counted p_i and p_j in different ways: While p_j includes every paper in our dataset, p_i includes only those papers published between 2009 and 2020, with "cleaned" reference lists of length > 0. Details regarding the reference list cleaning procedure can be found in the main text. We row-normalized N_{ij} when creating the directed citation network to account for differing total numbers of citations given by the set of papers in each journal.

We then performed a clustering analysis on this citation network, to determine journal (node) subsets whose intra-citation (edge weight) is dense and whose inter-citation (edge weight) is sparse. We used a generalized Louvain[9] community detection algorithm, freely available through the MATLAB Brain Connectivity Toolbox[10], to maximize a modularity quality index Q defined as[11]:

$$Q \equiv \frac{1}{s} \sum_{ij} \left[J_{ij} - P_{ij} \right] \delta(c_i, c_j). \tag{S1}$$

The scalar P_{ij} is the expected edge weight between nodes i and j in a suitably-chosen null model, $s = \sum_{ij} J_{ij}$ is the total edge weight of the network, and $\delta(c_i, c_j)$ is a Kronecker delta that is 1 if i and j are in the same community (*i.e.* if community indices $c_i = c_j$) and 0 otherwise. The modularity index Q can thus be thought of as the difference between the fraction of total edge weight that connects within-community nodes and that expected fraction under some null model. Many different null models are employed in the literature, each specific to the data and scientific question of interest[12]. Here, we used a simple null model for a directed network[13]:

$$P_{ij} = \frac{k_i^{out} k_j^{in}}{s}, \tag{S2}$$

where $k_i^{out} \equiv \sum_j J_{ij}$ is the weighted out-degree of node i, and $k_j^{in} \equiv \sum_i J_{ij}$ is the weighted in-degree of node j. The quantity k_i^{out}/s, the fraction of total edge weight out of node i, is the weighted probability that an edge chosen at random flows out of node i. Similarly, k_j^{in}/s, the fraction of total edge weight into node j, is the weighted probability that an edge chosen at random flows into node j. The quantity P_{ij}/s is thus equal to the joint weighted probability that, were an edge to be chosen at random, it would flow from i to j. Maximization of Q is accomplished by varying c, the partition of network nodes into communities. An important note is that, in practice, the non-symmetric modularity matrix $B_{ij} = \frac{1}{s}\left[J_{ij} - P_{ij}\right]$ is symmetrized before maximizing Q for algorithmic ease[13]; this operation does not change the scalar Q since $\sum_{ij} B_{ij} = \sum_{ij} B_{ji}$. We also note that the algorithm that maximizes Q is not guaranteed to find the global optimum[14], so we employed a modularity fine-tuning scheme in which Q maximization

was attempted and the resultant node partitioning fed back into the algorithm as an initial guess, resulting in a new maximal Q value and partitioning, until modularity ceased to increase.

The resultant clustering is shown in Fig. S4. Note that journals in the "high energy" half of the originally defined "high energy physics and astronomy" subfield (*Journal of High Energy Physics*, *Physical Review D*) are in a different community than the astronomical journals in that subfield (*Astrophysical Journal*, *Monthly Notices of the Royal Astronomical Society*, *Astronomy & Astrophysics*), with very limited inter-citation among those two halves. Furthermore, among other changes, the journal *Physics Letters B* has been clustered in the same community as the high energy physics journals, with strong inter-citation between these three journals. These observations motivated our final decisions regarding subfield delineation, shown in Fig. S5. We split the high energy journals from the astronomy and astrophysics journals, forming two separate subfields, and added *Physics Letters B* to the high energy subfield both because of its citation communication with the other journals in that subfield and to bolster the number of journals (3) in that subfield. We note that subfields could be defined in many ways, and the procedure just outlined is but one way of drawing boundaries in order to make observations on the level of subfield.

D. Alphabetical author ordering in high energy physics

We define the author gender category of papers according to the genders of those papers' first and last authors. The assumption we make with this definition is that author lists are ordered such that first and last author positions designate authors who are lead researchers on the paper and authors who are senior/supervisory researchers on the paper, respectively. This is not necessarily the case for papers in the high energy physics subfield, whose authors are often ordered alphabetically. In this section, we quantify the degree of alphabetical author ordering in the high energy physics subfield and explore its consequences for our analysis.

Fig. S6a shows the fraction of all multiple-author papers in each subfield whose first and last authors are in alphabetical order. This fraction is ≈ 0.875 in the high energy physics (HEP) subfield, well above the value it would have according to chance (0.5). Note that other subfields also other show alphabetical ordering at rates above chance, most notably the nuclear physics subfield; however, the HEP subfield shows significantly greater alphabetical ordering. Fig. S6b shows the fraction of all multiple-author papers in the HEP subfield whose first and last authors are in alphabetical order, grouped by publication year. Although this fraction has dropped slightly, from ≈ 0.878 in 1995 to ≈ 0.839 in 2020, it has remained well above 0.5 for every year covered in our data set. Finally, Fig. S6c shows the fraction of all multiple-author papers in the HEP subfield whose first and last authors are in alphabetical order, grouped by author gender category. Fractions are approximately equal for HEP papers written by MM vs. W||W teams, indicating that authors of papers classified as W||W are equally likely to be alphabetically ordered as authors of papers classified as MM.

Next, we explore the impact of alphabetical author ordering on our author gender categorization scheme. Suppose that papers have a "true" ordering that reflects author seniority or contribution, and that alphabetizing author names is a random reshuffling of this order. Let each paper have N total authors, and $n \leq N$ authors categorized as women. The probability that alphabetizing author names results in a paper given the author gender category W||W is the probability a woman is picked at random from the N authors for the first author position, plus the probability that a woman is picked at random from the N authors for the last author position, minus the probability that two women are picked at random from the N authors for both the first and last author position. The subtraction is necessary to avoid double-counting, since the first two events are not mutually exclusive. Then,

$$p_a(W||W) = \frac{2n}{N} - \frac{n(n-1)}{N(N-1)}, \tag{S3}$$

where $p_a(W||W)$ is the probability that an alphabetized author list is designated W||W. Thus, alphabetized papers in the HEP subfield are designated W||W with a probability that depends only on n/N, the fraction of authors in the author list that are women, and N, the size of the author list. As a result, the designation W||W for these papers reflects only the demographics of the author list, and not the contributions or seniority of women authors in the author list. Papers designated W||W in HEP might therefore be argued to constitute a broader class (of papers with randomly selected women co-authors) than those designated W||W in other subfields, for which women must be senior or highly contributing authors. The GAM should still be able to fit expected citation rates for this broader paper category, since it predicts author gender category according to publication journal and year, which correlate with n/N, and the log of paper team size, which directly encodes N. Thus, that demographic and citation imbalances still exist for this broader category of HEP W||W papers, as shown in the main text and throughout this supplement, is a striking result.

E. Coauthorship network determination

In order to study the dependence of over-/under-citation behavior on the scientific network of authors and papers, we quantified a co-author network for each paper i, consisting of all papers in our data set published the year of paper i's publication or earlier. Each node corresponds to a paper, and edges are drawn between nodes if the corresponding papers share at least one author. Note that edges exist between nodes if any author is shared between papers, not just first or last authors. Neighborhoods within this co-author network around paper i can then be defined to indicate sets of papers of varying proximity to paper i. For example, neighborhood $\mathcal{C}_1(i)$ consists of all papers of path length ≤ 1 from paper i, or all papers written by all of the co-authors of paper i (necessarily including paper i). Neighborhood $\mathcal{C}_2(i)$ consists of all papers of path length ≤ 2 from paper i, or all papers written by all of the co-authors of paper i and all of their co-authors. We could then quantify the over-/under-citation behavior of classes of papers as a function of the proximity of cited papers to each citing paper on its network. More specifically, for each paper i, we considered two separate subsets of its "cleaned" reference list (Section S2E): those cited papers within $\mathcal{C}_2(i)$ (thus more proximal to i), and those cited papers outside of $\mathcal{C}_2(i)$ (thus less proximal to i). We then pooled these separate sets of cited papers according to the author gender category of the citing papers, and calculated over-/under-citation of MM and W||W papers within these groups of more proximal and less proximal cited papers.

We also quantified homophily in the co-authorship network around each paper i, by defining the man-author over-representation ($\mathrm{MA}_{\mathrm{overrep}}$) in the network around paper i. We defined this measure to be the difference between the proportion of men within $\mathcal{C}_2(i)$ (excluding the co-authors of paper i) and the proportion of men in the entire field when paper i was published. Note that for all co-authors in $\mathcal{C}_2(i)$ who were neither the first nor last author of any paper in our data set, gender had not been algorithmically assigned. Thus, all proportions were taken only over those co-authors whose gender was assigned.

S4. SUPPLEMENTARY RESULTS

The following sections contain further details regarding the results presented in the main text. Each subsection header in the *Results* section in the main text has an identical corresponding header here for ease of cross-reference.

A. Time-varying demographics of published papers

Fig. 1 presents a demographic overview of the data set, and the main text describes general trends within that data. Note that although woman-authored (W||W) papers represent a small proportion of those tracked in Fig. 1a, the proportion of papers authored by first and last authors with names assigned to women (WW) remains significantly lower, growing from 1.7% in 1995 to only 3.6% in 2020. The fastest growing subset of woman-authored papers are those with assigned woman first authors and assigned man last authors (WM): the fraction of these papers grew from 0.06 in 1995 to 0.14 in 2020, for a growth rate of 229% over 25 years. The strong growth of this subset of papers might reflect the increasing number of junior woman scientists across physics. Other growth rates are 213% for MW papers and 210% for WW papers over 25 years.

Although the proportion of W||W papers increases in all subfields, individual journals vary significantly in that proportion and in its rate of change (Fig. 1b-c). Averaged over years, journals with the lowest fraction of W||W papers are *Reviews of Modern Physics* (0.11), *Journal of High Energy Physics* (0.15), and *Journal of Fluid Mechanics* (0.15). These three journals also have the lowest fraction of W||W papers published in 2020 among all journals, indicating that this trend is not merely due to their older age and the general increase of woman authors over time. Journals with the highest fraction of year-averaged W||W papers are *Nanoscale* (0.41), *Soft Matter* (0.40), and *ACS Applied Materials & Interfaces* (0.40).

B. Citation imbalance exists & varies by citing venue

Fig. 2 shows that (i) the citation behavior of the papers in our dataset is imbalanced with respect to the cited author gender category, and (ii) this imbalance varies according to citing subfield and journal. The main text discusses general trends in this behavior according to subfield. We note here that, in calculating citation gaps and trends, we consider only the citation behavior of papers published in 2009 or later, since these papers are more likely than those published earlier to cite other papers in our (1995 – 2020) dataset. To increase statistical power, the citation behavior reported in this section is aggregated over all papers published in 2009 or later, even those that could not be assigned to an author gender category.

As shown in finer detail in Fig. 2d, citation behavior at the journal level varies more widely still. Here, each point in the space of MM over-/under-citation versus W||W over-/under-citation shows the collective citation imbalance of papers grouped according to their publishing journal. The data follow an approximately linear trend with negative slope, indicating the correlation between over-citation of MM papers and under-citation of W||W papers. Journals that lie more deeply in the lower right quadrant host papers that collectively exhibit greater citation preference for MM papers, whereas journals that lie in the upper left quadrant host papers that collectively exhibit greater citation preference for W||W papers. Most journals are located in the lower right quadrant, indicating their preference toward MM over-citation, and only three journals (*Journal of Nuclear Materials*, *Astronomy & Astrophysics*, and *Soft Matter*) show a statistically significant preference for citing W||W papers.

C. Citation imbalance varies by citing actor

Fig. 3 demonstrates that citation behavior varies according to citing author gender category. The main text contains a concise description of this variance; here, we address it in greater detail. MM citing papers published in 2009 or later over-cite other MM papers by 2.05%, and under-cite W||W papers by 6.53%, for a gender citation gap of approximately 8.58% (Fig. 3a). By contrast, W||W citing papers published in 2009 or later over-cite other W||W papers by 3.56%, and under-cite MM papers by 1.38%, for a citation preference in favor of other W||W papers of approximately 4.94% (Fig. 3a).

These citation behaviors vary widely across subfield (Figs. 3c-d) and journal (Fig. S7). MM papers in the general physics subfield show the highest citation gap, 16.47%, in favor of other MM papers. MM papers in the nuclear and atomic/molecular/optical subfields show the next highest citation gaps, each above 15%. By contrast, W||W papers in the astronomy/astrophysics and soft matter/biophysics subfields exhibit a citation preference toward other W||W papers, resulting in citation gaps in favor of other W||W papers of 8.42% and 6.53%, respectively. Interestingly, citation preference in favor of other W||W papers does not exist for W||W citing papers in the general physics, atomic/molecular/optical, and condensed matter subfields. Rather, papers in these subfields show under-citation of other W||W papers and over-citation of MM papers. W||W papers grouped into the general physics subfield show the greatest citation gap (8.8%) in favor of MM papers.

D. Stable and growing trends in citation imbalance over time

Fig. 3 also shows trends in citation imbalance over time, grouped according to citing author gender category and citing subfield. The main text contains concise descriptions of these trends; here, we discuss in greater detail. The citation gap in favor of MM papers exhibited by MM citers is larger in 2020 than in 2009 (Fig. 3b, red lines), due to an overall increase in MM over-citation and W||W under-citation over time. The fraction of citations actually given by MM citers to MM (W||W) papers over time maintains a steady positive (negative) gap with respect to the fraction expected to be given according to our model (Fig. S8a). By contrast, W||W citers show a consistent citation preference in favor of other W||W papers with a markedly different trend over time (Fig. 3b, blue lines). For these citers, the citation gap in favor of other W||W papers decreases over time, in contrast to the increasing citation gap in favor of other MM papers shown by MM citers. The fraction of citations given by W||W citers to MM (W||W) papers over time shows a narrowing negative (positive) gap with respect to the fraction expected to be given according to our model (Fig. S8b). The data indicate an approach toward citation parity over time by W||W citers.

When grouped according to subfield, citation behavior shows a variety of different trends over time (Figs. 3e, S9, S10). Citation gaps between MM authored papers and W||W authored papers vary in magnitude according to subfield, but are relatively stagnant over time in many cases (within error bars). Two subfields that notably do not exhibit this behavior are the general physics and condensed matter subfields, for which the citation gap increases between 2009 and 2020, to 16.67% and 10.7% in 2020, respectively.

E. Citation imbalance varies by citation proximity

Fig. 4 and the accompanying discussion in the main text explore differences in citation behavior according to whether citations reference work with which authors are likely to be familiar. Here, we discuss these trends in greater detail. We first note that reference lists of citing papers contain a range of reference proportions that are deemed proximal or familiar according to the two definitions of proximity detailed in the main text (Fig. S11). We additionally note that precise definitions of subfields do not seem to affect overall conclusions regarding citation behavior for within-subfield vs. outside-subfield citations: Fig. S12 shows this citation behavior given the subfield

boundaries informed entirely by *post-hoc* clustering described in Section S3 C and illustrated in Fig. S4, and it is qualitatively and quantitatively similar to the behavior displayed in Fig. 4a in the main text.

With both definitions of proximity or familiarity, we find similar results regarding the total citation gap between MM and W||W authored papers: For the set of unfamiliar citations, the citation gap is larger, with greater over-citation of MM papers and greater under-citation of W||W papers (Fig. 4a-b, black symbols). Out-of-subfield MM papers are over-cited by 1.90%, and out-of-subfield W||W papers are under-cited by -7.36%, while these rates are 0.66% and -1.80% for within-subfield MM and W||W papers, respectively. Co-author-distant MM papers are over-cited by 1.29%, and co-author-distant W||W papers are under-cited by -4.34%, while these rates are 0.57% and -1.39% for co-author-proximal MM and W||W papers, respectively. For reference, we note that these values for all forms of citation, independent of familiarity, are MM over-citation of 1.06% and W||W under-citation of 3.17% (Fig. 2b).

The suppression of the "familiar citation gap" and the enhancement of the "unfamiliar citation gap" together arise from the cumulative effects of two very different citation behaviors according to citing author gender category. Across familiar citations, both MM (Fig. 4a-b, red symbols) and W||W (Fig. 4a-b, blue symbols) citing teams show enhanced citation preference for their respective author gender categories. That is, MM teams preferably cite MM papers, while W||W teams preferably cite W||W papers. We note that these competing effects may be partially explained by an overall homophilic enhancement in the local co-authorship network around each paper, revealed by higher man-author overrepresentation in co-authorship networks around MM papers and lower man-author overrepresentation in co-authorship networks around W||W papers (see *Supplementary Methods*, Fig. S13).

Across unfamiliar citations, however, MM and W||W teams differ in their citation behavior. For W||W teams, citation preference for W||W papers is approximately erased. By contrast, for MM teams, citation preference for MM papers is not erased. Instead, this preference is enhanced for out-of-subfield citations, and slightly reduced but still significant for co-author-distant citations. Thus, we find that W||W teams preferably cite familiar W||W papers and cite unfamiliar W||W and MM papers approximately equitably. MM teams over-cite familiar MM papers and under-cite familiar W||W papers, a trend that is even more pronounced for unfamiliar papers. The overall result is a smaller familiar citation gap, due to the competing citation behaviors of W||W and MM citing teams, and a larger unfamiliar citation gap in favor of MM papers, due to the MM citing papers' persistent citation preference for other MM papers even when citing unfamiliar references.

Note that, although citation imbalance across familiar citations is approximately equal and opposite for W||W and MM citing teams (for both definitions of familiarity), it is not distributed equally across W||W and MM cited papers. Rather, W||W teams exhibit a citation preference for W||W papers (vertical coordinate of the dark blue markers in Figs. 4a,b) that is about double that of MM teams for MM papers (horizontal coordinate of the dark red markers in Figs. 4a,b). By contrast, MM teams under-cite W||W papers (vertical coordinate of the dark red markers in Figs. 4a,b) at a rate that is about double the under-citation of MM papers by W||W teams (horizontal coordinate of the dark blue markers in Figs. 4a,b).

F. Additional correlates of citation imbalance

Fig. 5 and the accompanying discussion in the main text investigate trends in citation behavior according to two additional correlates: the relative proportion of W||W published papers in each journal, and the reference list length of citing papers. Here, we provide more detail about both analyses. For the former, we observe that the 35 journals investigated in this paper generally show an increasing time-aggregated citation preference for W||W papers with an increasing time-aggregated fraction of W||W papers published (Fig. 5a). Correspondingly, the time-aggregated citation preference for MM papers decreases with an increasing time-aggregated fraction of W||W papers published (Fig. S14).

For individual citing papers, we observe a collective effect in which papers with longer reference lists tend to exhibit increased citation preference for W||W papers (Figs. 5b, S15). This trend holds independently for both MM and W||W citing teams, despite the fact that W||W papers in our dataset contain longer reference lists in comparison with MM papers (Fig. S16). Notably, the trend also remains stable over time, despite the fact that reference list length within our dataset generally increases over time (Fig. S17). More specifically, we group reference lists according to citing author gender and publishing year, and find that slopes of linear fits of MM over/under-citation as a function of reference list length are consistently negative and similar in magnitude for MM and W||W citing teams over publishing years between 2009 and 2020. Similarly, slopes of linear fits of W||W over/under-citation as a function of reference list length are consistently positive, stable over time, and almost identical in magnitude for MM and W||W citing teams (Fig. S18). Effectively, MM (W||W) citing teams show a reduction of approximately 0.43% (0.7%) MM over/under-citation for every 10 citations added to their reference lists. Both citing teams show an increase of approximately 1.5% W||W over/under-citation for every 10 citations added. Because the coefficient of determination of each linear fit is low (Fig. S19) due to the high variance of our data, the relation should be considered a weak but significant

collective effect rather than an exact measure of correlation for any particular paper subset.

We note that to eliminate potentially spurious effects of papers whose reference lists contain few within-database citations for which we could categorize author gender, we only performed our linear fitting procedure on papers with reference lists containing 20% or more within-database and author-gender-categorized citations. Results remain consistent if this threshold percentage is doubled to 40%, and results are qualitatively similar even if this threshold is removed (Fig. S20).

G. Citation imbalance of sub-categories of W||W papers

Sub-categories of papers written by at least one first or last author whose name was assigned "woman" were combined into the larger W||W category in the main text (Fig. 1a). This combination was necessary for statistical purposes: The demographic realities of the field of physics meant that papers written by the WM, MW, and WW teams on their own formed too-small subsets of our dataset with noisy citation statistics. However, examination of these sub-categories of W||W papers is quite useful, and would ideally be included in any full analysis, to distinguish between papers with women in junior vs. senior author positions. Here, we explore the citation behavior of these W||W sub-categories.

Fig. S21a shows the over-/under-citation behavior of all sub-categories of W||W papers, calculated over all papers in each sub-category published between 2009 and 2020. The over-/under-citation behavior of the larger W||W citer category is also shown for comparison with each sub-category. Papers with names assigned to men in the first or last author position, the WM and MW sub-categories, show lesser over-citation of W||W papers, and lesser under-citation of MM papers, than the broader W||W citer category. Of these two sub-categories, the WM sub-category shows larger over-citation of W||W papers and under-citation of MM papers. By contrast, papers with names assigned to women in the first and last author position, the WW sub-category, show much larger over-citation of W||W papers, and larger under-citation of MM papers, than the broader W||W citer category. Taken together, the data imply that WW teams especially drive the overall citation behavior of the W||W citer category, and that the effects of women first ("junior") authors are especially correlated with the over-citation of W||W papers and the under-citation of MM papers.

Figs. S21b-f show over-/under-citation behavior of all sub-categories of W||W papers over time. As in panel a, the over-/under-citation behavior of the larger W||W citer category is also shown for comparison with each sub-category. The data show similar trends to panel a: The over-/under-citation behavior of WM and MW teams is less favorable toward W||W papers than the behavior of the broader W||W citer group, with MW citers showing less favoritism toward W||W papers than WM citers. By contrast, the over-/under-citation behavior of WW papers is even more favorable toward W||W papers than the behavior of the broader W||W citer group. Importantly, for both WM and WW citing groups, this favoritism toward W||W papers decreases in magnitude from 2009 to 2020, in a similar manner to the trend over time exhibited by the broader W||W citer category. Note that the WU and UW sub-categories (papers in which one of the first or last author's names was assigned to the woman gender category, and the other name could not be assigned to a gender category) are shown for completeness in panels a, e, and f, but show noisy statistics and cannot be clearly interpreted in terms of assigned author gender.

H. Effects of self-citation on citation behavior

In the main results presented in this paper, self-citations were removed from all reference lists before performing any analyses of citation behavior. We made this choice because the phenomenon of self-citation is quite different than the phenomenon quantified in this paper, regarding how the gender perception of others influences authors' engagement with the rest of their field. Self-citation has been studied elsewhere in great detail [e.g., see Ref. 15], and its inclusion in our analyses would necessarily enhance gender homophily in citation behavior due to our definitions of author gender category and self-citation (see Section S2E). In this section, we explore the effects of including self-citations in our analyses.

We first note that MM teams and W||W teams self-cite at different rates. Fig. S22a shows overlaid histograms of the number of self-citations in all MM and W||W papers in our data set. These distributions each have medians of 1, but the mean of the MM distribution is larger (≈ 1.57) than that of the W||W distribution (≈ 1.40). Moreover, a Wilcoxon rank sum test (or Mann-Whitney U test) of the alternative hypothesis that the MM self-citation distribution is shifted to the right of the W||W self-citation distribution has a p-value $< 2.2\text{e-}16$, indicating that indeed the MM distribution is shifted toward higher self-citation numbers than the W||W distribution. This is an especially striking result in the context of Fig. S16, which shows that W||W citing papers generally contain longer reference lists than

MM citing papers. In other words, MM teams cite fewer papers in general than W||W teams, and still cite themselves more, in general, than W||W teams.

These self-citations, however, do not change the overall patterns of over-citation of MM papers and under-citation of W||W papers. Fig. S22b shows the over-/under-citation of MM and W||W papers by all citing papers in the data set published between 2009 and 2020, both excluding self-citations from reference lists (dotted lines) and including self-citations in reference lists (colored bars). Overall over-citation of MM papers is approximately the same both including and excluding self-citations, and the under-citation of W||W papers is slightly larger when including self-citations in reference lists.

When over-/under-citation is calculated separately according to citing author gender category, the homophilic effects of including self-citations in the analysis become much more obvious. Fig. S22c shows over-/under-citation of MM and W||W papers, for all MM papers in the data set published between 2009 and 2020 (red) and all W||W papers in the data set published between 2009 and 2020 (blue). Over-/under-citation measures both excluding self-citations from reference lists (dotted lines) and including self-citations in reference lists (colored bars) are shown. The inclusion of self-citations in MM papers enhances their over-citation of MM papers and under-citation of W||W papers, and the inclusion of self-citations in W||W papers enhances their over-citation of W||W papers and under-citation of MM papers. These homophilic effects compete with each other, however, leading to the much smaller effects in overall over-/under-citation shown in Fig. S22b. We note that in calculating all over-/under-citation measures shown in Figs. S22b and c, we cleaned each reference list in an identical manner to that discussed in Section S2E, with the only difference being that we either explicitly excluded self-citations (as in the primary analyses) or included them.

I. Co-authorship neighborhoods differ according to citing author gender category

In the main text of the paper, we calculated differences in citation behavior when authors were citing within their local co-authorship neighborhoods vs. outside of those neighborhoods. Those analyses found that both MM and W||W teams tend to exercise significant homophily when citing within their local co-authorship neighborhoods, with MM papers over-citing other MM papers and W||W papers over-citing other W||W papers. By contrast, when citing outside their local co-authorship neighborhoods, W||W teams tend to cite approximately equitably, while MM teams tend to continue over-citing other MM papers and under-citing W||W papers. In this section, we will explore these co-authorship neighborhoods in more depth.

Figs. S23a and b show distributions of co-authorship neighborhood size at two different scales, $|\mathcal{C}_1|$ and $|\mathcal{C}_2|$ (defined in Section S3E), for MM vs. W||W papers. Results show that these local neighborhood sizes tend to be larger around W||W papers in comparison to MM papers. The median of $|\mathcal{C}_1|$ around W||W papers is 38, larger than the median of $|\mathcal{C}_1|$ around MM papers (32). A Wilcoxon rank sum test (or Mann-Whitney U test) of the alternative hypothesis that the MM distribution is shifted with respect to the W||W distribution has a p-value $<$ 2.2e-16, indicating that indeed the MM distribution is shifted toward lower $|\mathcal{C}_1|$ values with respect to the W||W distribution. Similarly, the median of $|\mathcal{C}_2|$ around W||W papers is 974, larger than the median of $|\mathcal{C}_2|$ around MM papers (587). A Wilcoxon rank sum test of the alternative hypothesis that the MM distribution is shifted with respect to the W||W distribution also has a p-value $<$ 2.2e-16. These trends also hold when considering only the neighborhoods around papers with a limited number of direct co-authors (< 20), which removes the influence of giant collaboration papers (Figs. S24a,b). For these papers, the median of $|\mathcal{C}_1|$ around W||W papers is 37, while that around MM papers is 32. The median of $|\mathcal{C}_2|$ around W||W papers is 932, while that around MM papers is 577. Wilcoxon rank sum tests indicate that the distributions of $|\mathcal{C}_1|$ and $|\mathcal{C}_2|$ around MM papers are shifted with respect to their W||W counterparts ($p <$ 2.2e-16 in all cases). These results likely are a consequence of the fact that W||W papers tend to have larger numbers of direct co-authors than MM papers. The distribution of the number of co-authors on W||W papers in the data set has a median of 4, while the median number of co-authors on MM papers in the data set is 3. A Wilcoxon rank sum test indicates in this case also that the MM distribution is shifted toward lower co-author numbers with respect to the W||W distribution ($p <$ 2.2e-16). This phenomenon can also be seen in Fig. S2d: The ratio of W||W papers to MM papers trends toward higher values as the log of paper team size, or number of direct co-authors, increases.

Next, we analyze the gender makeup of local co-authorship neighborhoods around MM and W||W papers. Figs. S23c and d show distributions of the fraction of MM papers, $p(MM)$, in the $\mathcal{C}_1$ and $\mathcal{C}_2$ neighborhoods surrounding MM vs. W||W papers. Fractions are taken with respect to only those papers in each local neighborhood whose author gender category is known: $p(MM) \in \mathcal{C}_n$ is the number of MM papers in $\mathcal{C}_n$ divided by the total number of MM and W||W papers in $\mathcal{C}_n$ around each paper. Results show that MM papers contain a higher fraction of MM papers in both $\mathcal{C}_1$ and $\mathcal{C}_2$ in comparison to W||W papers. The median of $p(MM) \in \mathcal{C}_1$ around MM papers is 0.90, while that around W||W papers is 0.54. Wilcoxon rank sum tests indicate that these distributions are shifted with respect to each other ($p <$ 2.2e-16). The median of $p(MM) \in \mathcal{C}_2$ around MM papers is 0.80, while that around W||W papers is 0.69. Wilcoxon rank sum tests also indicate that these distributions are shifted with respect to each other ($p <$

2.2e-16). These trends also hold when considering only the neighborhoods around papers with a limited number of direct co-authors (Figs. S24c,d). For these papers, the median of $p(MM) \in \mathcal{C}_1$ around MM papers is 0.90, while that around W||W papers is 0.54. The median of $p(MM) \in \mathcal{C}_2$ around MM papers is 0.80, while that around W||W papers is 0.70. Wilcoxon rank sum tests indicate that the distributions of $p(MM) \in \mathcal{C}_1$ and $p(MM) \in \mathcal{C}_2$ around MM papers are shifted with respect to their W||W counterparts ($p <$ 2.2e-16 in all cases).

We note that the gender makeup of local co-authorship neighborhoods could influence citation behavior when authors are citing within their local co-authorship neighborhoods vs. outside of those neighborhoods. If we assume that (i) expected citation rates of MM and W||W papers are not significantly different for papers within each $\mathcal{C}_2$ neighborhood vs. outside of it, and (ii) citations of each MM paper within its $\mathcal{C}_2$ neighborhood are a random sampling of that neighborhood, we can conclude that the homophilic enhancement of MM papers in $\mathcal{C}_2$ neighborhoods around MM papers leads to their greater over-citation of MM papers. The same would apply to homophilic enhancement of W||W papers in $\mathcal{C}_2$ neighborhoods around W||W papers, and their greater over-citation of W||W papers. This homophilic enhancement makes the contrast in the out-of-$\mathcal{C}_2$ over-/under-citation behavior of MM vs. W||W teams (Fig. 4b) even more striking. For citation subsets outside of their $\mathcal{C}_2$ neighborhoods, with presumably lower homophilic enhancement, W||W teams cite near equity, while MM teams maintain a significant citation preference for other MM papers. Thus, even for papers that are more distant in their co-authorship network, MM papers maintain their homophilic preference.

S5. SUPPLEMENTARY DISCUSSION

A. The citation diversity statement in other journals

Papers with citation diversity statements (CDSs) have now been published in at least 33 different journals. These include discipline general journals such as *Proceedings of the National Academy of Sciences*, *Science Advances*, *Nature Communications*, and *Scientific Reports*; physics journals (*New Journal of Physics*), discipline specific journals in the Nature family (e.g., *Nature Machine Intelligence*, *Nature Biomedical Engineering*, *Nature Reviews Neuroscience*, *Nature Neuroscience*, *Communications Biology*); and other discipline specific journals especially in the biological sciences: *ACS Catalysis*, *Annals of the International Communication Association*, *Annual Review of Vision Science*, *Biological Psychiatry*, *Brain and Language*, *Cerebral Cortex*, *Current Opinion in Behavioral Sciences*, *Current Opinion in Neurobiology*, *Developmental Cognitive Neuroscience*, *Developmental Psychobiology*, *Handbook of the Mathematics of the Arts and Sciences*, *Human Brain Mapping*, *Journal of Neural Engineering*, *Journal of Neuroscience*, *Journal of Tissue Engineering*, *Journal of Vision*, *Network Neuroscience*, *Neuroimage*, *Neuropsychopharmacology*, *The Gerontologist*, *Translational Psychiatry*, *Trends in Cognitive Sciences*, and *Visual Cognition*.

REFERENCES

[1] Carl T. Bergstrom, Jevin D. West, and Marc A. Wiseman. The EigenfactorTM Metrics. *Journal of Neuroscience*, 28(45):11433–11434, nov 2008.

[2] Csaba Toth, Elizabeth Durham, Murat Kantarcioglu, Yuan Xue, and Bradley Malin. SOEMPI: A Secure Open Enterprise Master Patient Index Software Toolkit for Private Record Linkage. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2014:1105–1114, 2014.

[3] Gender API, 2021.

[4] Michelle L. Dion, Jane Lawrence Sumner, and Sara Mc Laughlin Mitchell. Gendered Citation Patterns across Political Science and Social Science Methodology Fields. *Political Analysis*, 26(3):312–327, 2018.

[5] Jane Pilcher. Names and "Doing Gender": How Forenames and Surnames Contribute to Gender Identities, Difference, and Inequalities. *Sex Roles*, 77(11-12):812–822, 2017.

[6] Corinne A. Moss-Racusin, John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41):16474–16479, 2012.

[7] Jordan D. Dworkin, Kristin A. Linn, Erin G. Teich, Perry Zurn, Russell T. Shinohara, and Danielle S. Bassett. The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience*, 23(8):918–926, aug 2020.

[8] Simon N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edition, 2017.

[9] Vincent D. Blondel, Jean Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 2008.

[10] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.

[11] Mark E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 69(2 2):1–15, 2004.

[12] Danielle S. Bassett, Mason A. Porter, Nicholas F. Wymbs, Scott T. Grafton, Jean M. Carlson, and Peter J. Mucha. Robust detection of dynamic community structure in networks. *Chaos*, 23(1), mar 2013.

[13]Elizabeth A. Leicht and Mark E.J. Newman. Community structure in directed networks. *Physical Review Letters*, 100(11):1–4, 2008.
[14]Benjamin H. Good, Yves-Alexandre de Montjoye, and Aaron Clauset. Performance of modularity maximization in practical contexts. *Physical Review E*, 81(4):046106, apr 2010.
[15]Molly M. King, Carl T. Bergstrom, Shelley J. Correll, Jennifer Jacquet, and Jevin D. West. Men Set Their Own Cites High: Gender and Self-citation across Fields and over Time. *Socius: Sociological Research for a Dynamic World*, 3:237802311773890, 2017.

LIST OF SUPPLEMENTARY TABLES AND FIGURES

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Appl Optics	934	860	1201	1054	903	809	777	901	698	781	849	1058	982	980
Nat Photonics	0	0	0	0	0	0	0	0	0	0	0	0	56	74
Opt Express	0	0	57	128	91	94	164	198	437	764	1231	1408	1960	2322
Opt Lett	848	674	590	607	585	556	580	622	728	836	985	1066	1146	1006
Phys Rev A	1012	1264	1254	1276	1308	1381	1527	1828	1594	1700	2039	2088	2259	2502
Phys Rev B	3484	4708	4462	4383	4591	4754	4722	5473	4694	4964	4350	1904	5744	5784
Adv Mater	130	193	203	268	285	367	374	402	420	414	581	565	734	773
Phys Chem Chem Phys	0	0	0	0	745	775	740	764	723	677	462	577	591	755
Adv Funct Mater	0	0	0	0	0	0	64	104	130	148	255	279	461	422
Nat Mater	0	0	0	0	0	0	0	36	119	132	139	139	133	129
Nat Phys	0	0	0	0	0	0	0	0	0	0	26	114	132	147
Phys Rev Lett	2428	2684	2678	3048	2854	3046	2994	2962	2962	3575	3693	3760	3546	3907
Phys Rev X	0	0	0	0	0	0	0	0	0	0	0	0	0	0
New J Phys	0	0	0	2	21	32	24	101	160	203	258	331	452	682
Rev Mod Phys	26	27	29	35	97	29	32	32	38	29	30	32	33	36
Phys Rev D	1267	1488	1627	1728	1792	2022	1955	2274	1963	2277	2246	2372	2267	2862
Astrophys J	2095	2124	2221	2205	2196	2362	2516	2299	2435	2473	2595	2790	2796	2129
Mon Not R Astron Soc	670	716	751	884	859	897	1010	1050	1139	1222	1316	1553	1490	1652
Astron Astrophys	1012	1258	1344	1320	1208	1412	1814	1821	1936	1870	1879	1935	1978	1789
J High Energy Phys	0	0	0	220	353	532	606	732	804	883	857	1024	1247	1281
ACS Nano	0	0	0	0	0	0	0	0	0	0	0	0	53	296
Appl Phys Lett	2418	2560	2366	2351	2591	2638	2749	3148	3257	3730	4414	6152	5817	5450
ACS App Mater Inter	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nano Lett	0	0	0	0	0	0	151	294	347	459	490	555	689	818
Nanoscale	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J Nucl Mater	142	188	144	136	178	201	162	160	143	192	200	233	77	359
Nucl Fusion	95	112	130	118	110	101	177	138	161	146	189	90	200	100
Nucl Instr A	670	704	610	561	654	484	416	756	448	500	684	538	555	582
Phys Rev C	645	791	774	867	864	832	797	926	813	881	852	861	934	904
Phys Lett B	1569	1654	1546	1744	1430	1375	1335	1155	968	1035	955	999	839	930
Biophys J	532	618	571	588	605	582	578	615	734	747	799	906	860	1040
J Comput Phys	204	236	233	229	228	229	265	239	276	306	308	391	522	413
J Fluid Mech	389	379	370	334	340	321	388	398	379	377	400	486	498	450
Phys Rev E	1240	1683	1872	1930	1888	2033	2325	2638	2273	2282	2525	2330	2255	2361
Soft Matter	0	0	0	0	0	0	0	0	0	0	46	109	171	296
SUM	21810	24921	25033	26016	26776	27864	29242	32066	30779	33603	35653	36645	41477	43231

TABLE S1. The number of papers contained in our data set, grouped by journal and year. Row and column sums are shown to summarize. The table is continued on the following page for the years 2009-2020.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	SUM
Appl Optics	913	995	1058	1185	1187	1311	1548	1426	1532	1520	1414	910	27786
Nat Photonics	82	105	96	120	139	130	113	121	109	103	115	71	1434
Opt Express	2548	2945	2983	3173	3288	3335	3321	2907	3061	3063	3303	1929	44710
Opt Lett	1309	1408	1605	1755	1618	1779	1478	1465	1327	1525	1514	1232	28844
Phys Rev A	2536	2858	2722	2759	2785	2660	2545	2729	2644	2619	2476	1020	53385
Phys Rev B	5676	6047	6136	5647	4774	4813	4892	5376	5396	5099	5025	2182	125080
Adv Mater	665	777	789	867	872	967	988	1152	1349	1504	1304	593	17536
Phys Chem Chem Phys	1247	1712	2314	1804	2265	2878	3433	3500	3304	3015	2617	1147	36045
Adv Funct Mater	442	481	533	569	636	813	770	873	806	1369	1382	805	11342
Nat Mater	134	137	134	141	151	153	169	172	162	146	162	80	2568
Nat Phys	146	158	163	137	122	131	153	177	187	188	190	84	2255
Phys Rev Lett	3414	3105	3246	3787	3557	2787	2502	2345	2492	2782	2644	1215	78013
Phys Rev X	0	0	38	70	92	216	173	198	220	277	235	121	1640
New J Phys	806	812	733	857	915	819	854	748	576	674	642	217	10919
Rev Mod Phys	46	73	40	45	45	34	38	41	40	41	38	10	996
Phys Rev D	2813	2932	2988	3333	3230	3468	3373	3507	3376	3535	3677	1669	66041
Astrophys J	2796	2501	2473	3075	2889	2787	3008	3006	3075	2969	3162	2044	67021
Mon Not R Astron Soc	1774	1970	2398	2616	2686	2851	3097	3209	3486	3850	4001	2434	49581
Astron Astrophys	1786	1917	1939	1892	1807	1737	1778	1831	1782	1889	2017	1260	44211
J High Energy Phys	1270	1416	1651	1869	1948	2008	2172	2128	1950	2138	2191	898	30178
ACS Nano	497	902	1141	1191	1178	1328	1268	1250	1325	1290	1379	834	13932
Appl Phys Lett	4675	4456	4419	4977	5362	5041	3436	3044	2763	2286	1898	1398	93396
ACS App Mater Inter	400	516	665	952	1778	2761	3349	4055	4862	4889	5182	3879	33288
Nano Lett	804	855	955	1078	996	1103	1260	1169	1136	1105	1160	508	15932
Nanoscale	46	360	653	1015	1546	1840	2259	2174	2084	2441	2343	1667	18428
J Nucl Mater	318	338	404	416	638	633	679	548	552	699	553	403	8696
Nucl Fusion	277	77	318	187	326	193	335	204	526	400	458	179	5347
Nucl Instr A	941	555	1027	625	741	756	737	997	644	799	1037	618	17639
Phys Rev C	1048	1013	1080	1120	1101	1081	1069	1057	1041	1015	989	359	23714
Phys Lett B	928	769	1010	869	778	817	825	905	898	867	836	257	27293
Biophys J	832	825	696	601	562	577	541	515	508	496	415	340	16683
J Comput Phys	483	494	468	425	629	675	696	713	718	647	667	531	11225
J Fluid Mech	450	524	565	557	684	687	670	740	814	852	1021	790	13863
Phys Rev E	2456	2310	2506	2459	2503	2393	2467	2316	2263	2095	2027	836	56266
Soft Matter	590	710	1353	1358	1200	966	943	967	907	965	958	450	11989
SUM	45148	47053	51299	53531	55028	56528	56939	57565	57915	59152	59032	32970	1067276

	Estimate	Std. Error	p-value	
as.factor(journal)acs nano	-0.10985	0.02527	1.38E-05	***
as.factor(journal)advanced functional materials	-0.09629	0.02811	6.12E-04	***
as.factor(journal)advanced materials	-0.27828	0.02608	<2E-16	***
as.factor(journal)applied optics	-0.63087	0.02233	<2E-16	***
as.factor(journal)applied physics letters	-0.51038	0.01828	<2E-16	***
as.factor(journal)astronomy & astrophysics	0.39037	0.01986	<2E-16	***
as.factor(journal)astrophysical journal	0.08596	0.01777	1.31E-06	***
as.factor(journal)biophysical journal	-0.08018	0.02342	6.18E-04	***
as.factor(journal)journal of computational physics	-0.87401	0.03420	<2E-16	***
as.factor(journal)journal of fluid mechanics	-0.84101	0.03329	<2E-16	***
as.factor(journal)journal of high energy physics	-0.71408	0.02348	<2E-16	***
as.factor(journal)journal of nuclear materials	-0.48532	0.03824	<2E-16	***
as.factor(journal)monthly notices of the royal astronomical society	0.15656	0.01862	<2E-16	***
as.factor(journal)nano letters	-0.27602	0.02469	<2E-16	***
as.factor(journal)nanoscale	0.05094	0.02351	3.02E-02	*
as.factor(journal)nature materials	-0.54199	0.05425	<2E-16	***
as.factor(journal)nature photonics	-0.89993	0.08120	<2E-16	***
as.factor(journal)nature physics	-0.84656	0.06666	<2E-16	***
as.factor(journal)new journal of physics	-0.52994	0.03292	<2E-16	***
as.factor(journal)nuclear fusion	-1.00783	0.08267	<2E-16	***
as.factor(journal)nuclear instruments & methods in physics research section a	-0.63845	0.03422	<2E-16	***
as.factor(journal)optics express	-0.59868	0.02021	<2E-16	***
as.factor(journal)optics letters	-0.71860	0.02302	<2E-16	***
as.factor(journal)physical chemistry chemical physics	-0.11128	0.01915	6.25E-09	***
as.factor(journal)physical review a	-0.53199	0.02054	<2E-16	***
as.factor(journal)physical review b	-0.54022	0.01754	<2E-16	***
as.factor(journal)physical review c	-0.09881	0.02947	8.00E-04	***
as.factor(journal)physical review d	-0.51881	0.01955	<2E-16	***
as.factor(journal)physical review e	-0.55850	0.01999	<2E-16	***
as.factor(journal)physical review letters	-0.72780	0.01973	<2E-16	***
as.factor(journal)physical review x	-0.87640	0.07716	<2E-16	***
as.factor(journal)physics letters b	-0.57424	0.02823	<2E-16	***
as.factor(journal)reviews of modern physics	-0.94860	0.12025	3.05E-15	***
as.factor(journal)soft matter	0.15263	0.02556	2.36E-09	***
as.factor(review)TRUE	0.08692	0.03100	5.05E-03	**

TABLE S2. Coefficients for linear terms in the GAM, errors of those terms, and p-values for the null hypothesis that each linear term is zero.

	edf	p-value	
s(month_from_base)	7.219	<2E-16	***
s(masked_seniority)	8.141	<2E-16	***
s(log_teamsize)	8.984	<2E-16	***

TABLE S3. Effective degrees of freedom of all smoothed terms in the GAM, and p-values for the null hypothesis that each smooth term is zero.

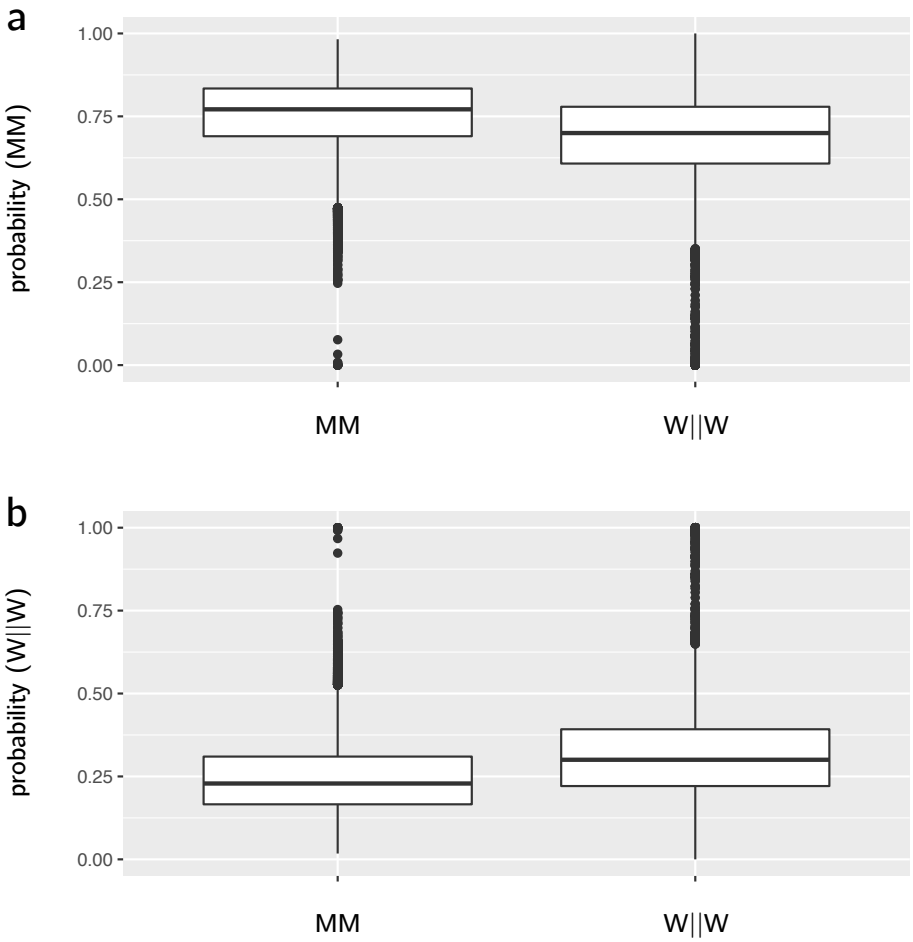

FIG. S1. **GAM-predicted author gender categories correlate with observed author gender categories.** (a) GAM-predicted probabilities that papers are written by MM teams, shown as box plots for all papers actually written by MM teams and all papers actually written by W||W teams. (b) GAM-predicted probabilities that papers are written by W||W teams, shown as box plots for all papers actually written by MM teams and all papers actually written by W||W teams. In both panels, outliers are drawn individually when they exist outside 1.5 times the inter-quartile range of the distribution.

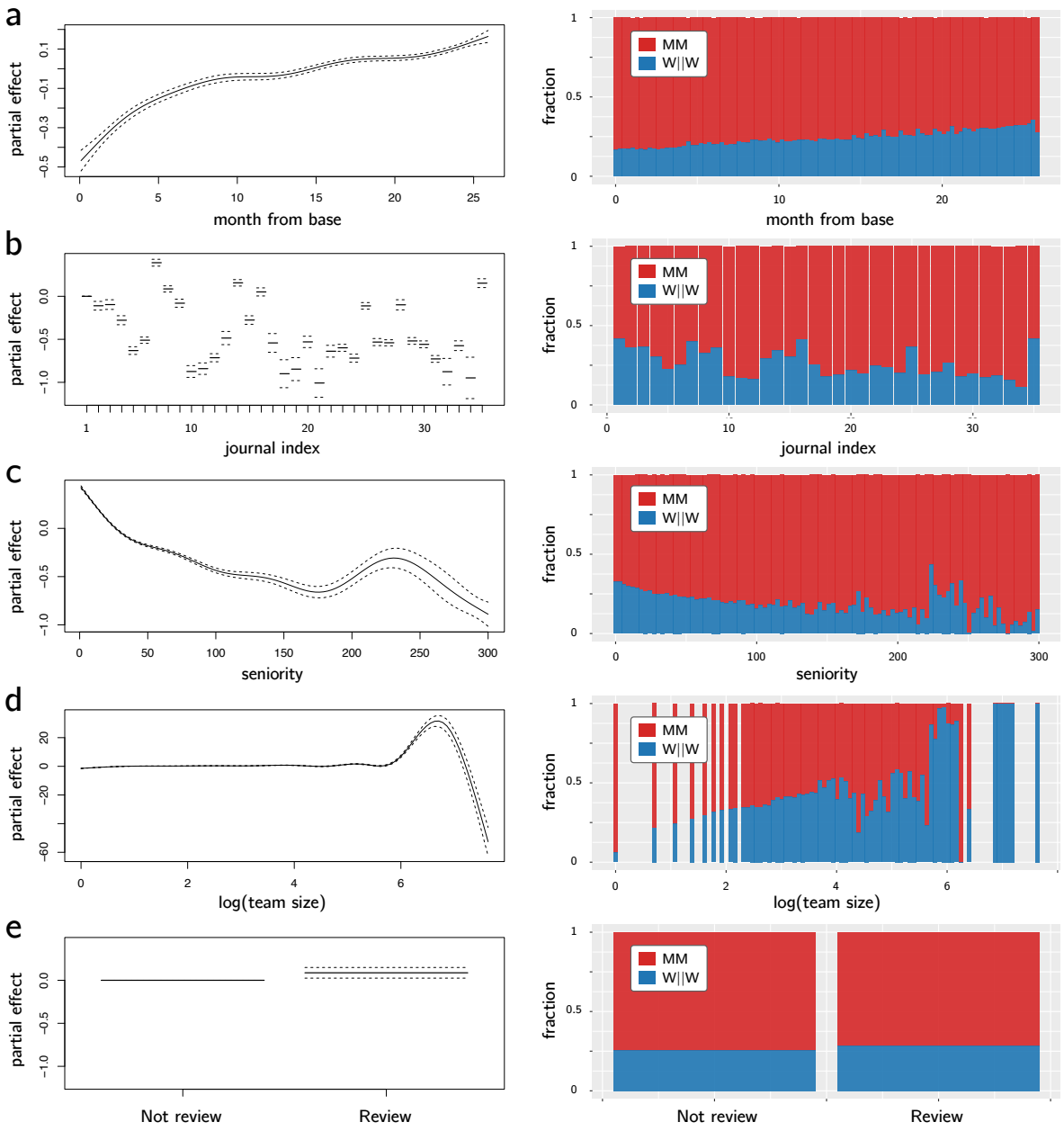

FIG. S2. **GAM-predicted variation in author gender category as a function of paper characteristics.** Panels show partial effect plots (left) and actual fractions of papers written by MM and W||W teams (right) for all paper characteristics included in the GAM. The paper characteristics are (a) publication month from base, or number of months (divided by 12 such that axis labels are in units of years) from the earliest publication date in the data set, (b) publishing journal, (c) seniority, or first and last authors' combined number of papers in the dataset, (d) log of team size, or total number of authors, and (e) categorization as non-review or review article. Partial effect plots show the component effect of each paper characteristic on the overall prediction that papers are written by W||W teams with respect to MM teams. Error bars show the 95% confidence interval for each effect. Journals are indexed by 1. ACS App Mater Interfaces, 2. ACS Nano, 3. Adv Funct Mater, 4. Adv Mater, 5. Appl Optics, 6. Appl Phys Lett, 7. Astron Astrophys, 8. Astrophys J, 9. Biophys J, 10. J Comput Phys, 11. J Fluid Mech , 12. J High Energy Phys, 13. J Nucl Mater, 14. Mon Not R Astron Soc, 15. Nano Lett, 16. Nanoscale, 17. Nat Mater, 18. Nat Photonics, 19. Nat Phys, 20. New J Phys, 21. Nucl Fusion, 22. Nucl Instrum Methods Phys Res A, 23. Opt Express, 24. Opt Lett, 25. Phys Chem Chem Phys, 26. Phys Rev A, 27. Phys Rev B, 28. Phys Rev C, 29. Phys Rev D, 30. Phys Rev E, 31. Phys Rev Lett, 32. Phys Rev X, 33. Phys Lett B, 34. Rev Mod Phys, 35. Soft Matter.

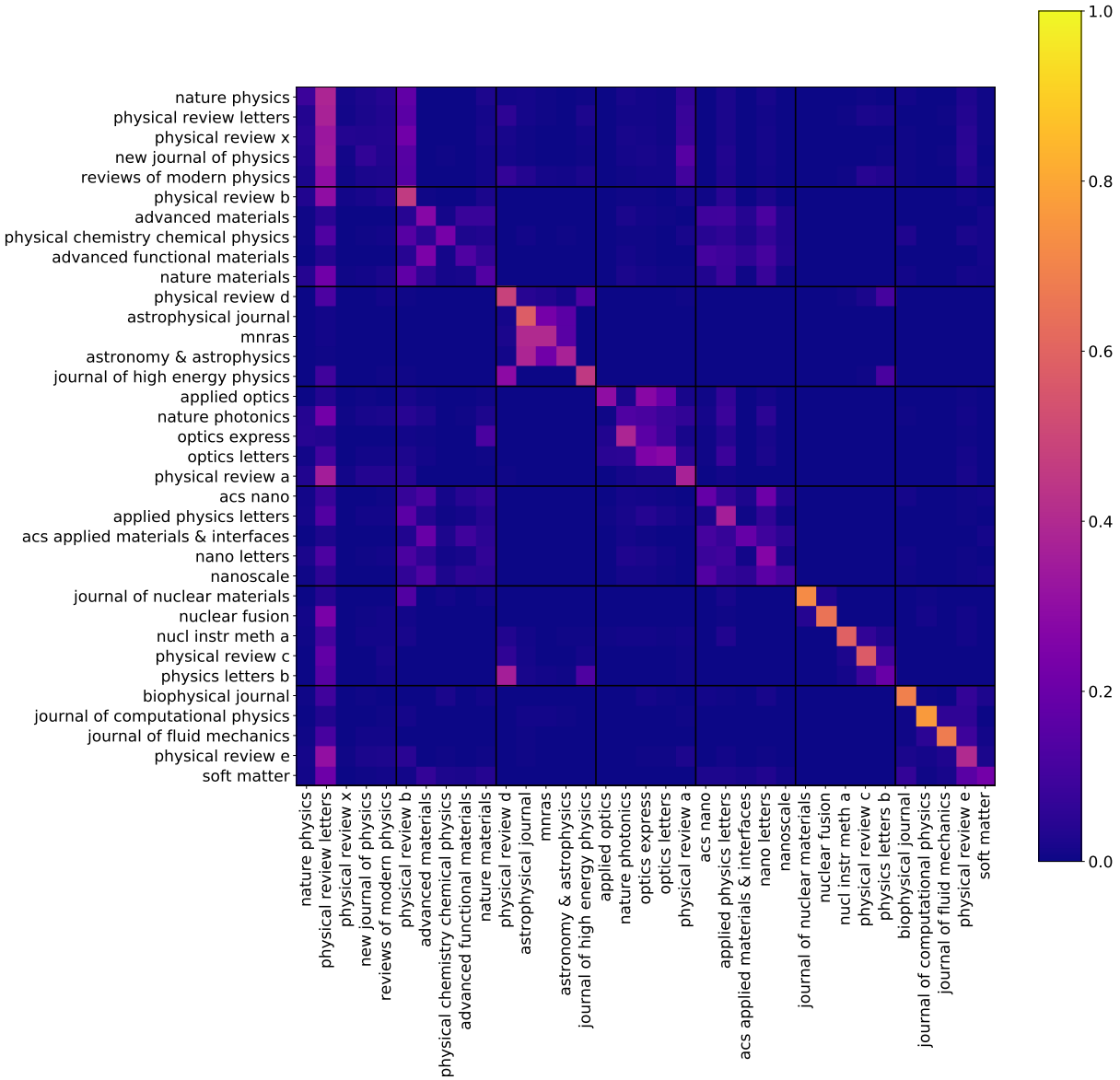

FIG. S3. **Directed citation network with initial subfield boundaries.** Each matrix element J_{ij} is colored according to the fraction of citations given by papers in citer journal i to papers in cited journal j, as indicated by the color bar. Details regarding the calculation of J_{ij} can be found in the supplementary text. Black lines mark initial subfield boundaries delineated according to pre-defined journal categories culled from the breakdown of the *Physical Review* family of journals and *Web of Science* journal categories.

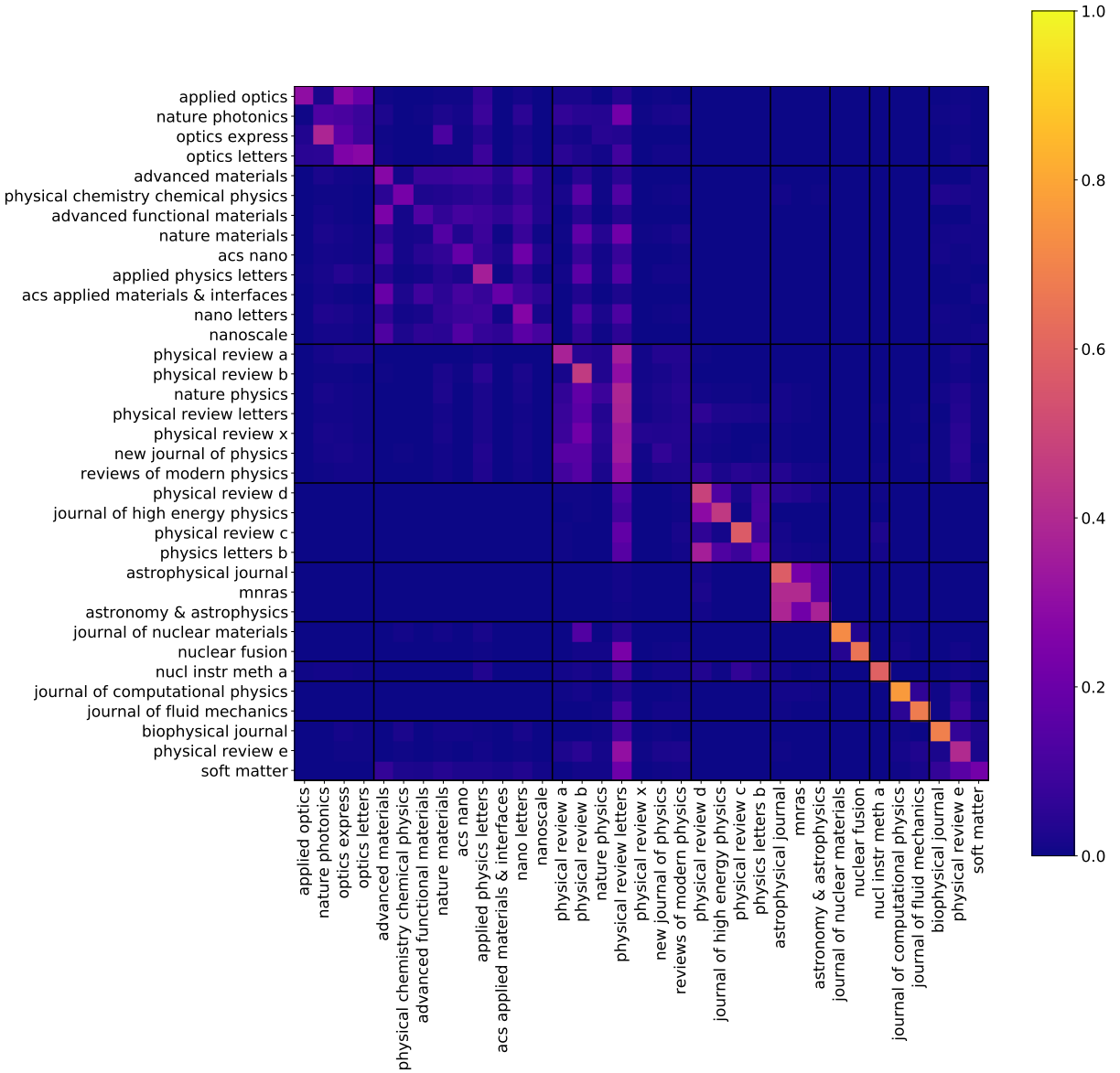

FIG. S4. **Citation network communities.** Each matrix element J_{ij} is colored according to the fraction of citations given by papers in citer journal i to papers in cited journal j, as indicated by the color bar. Journals have been grouped into communities, separated by black lines, that maximize the citation network modularity Q defined in the supplementary text.

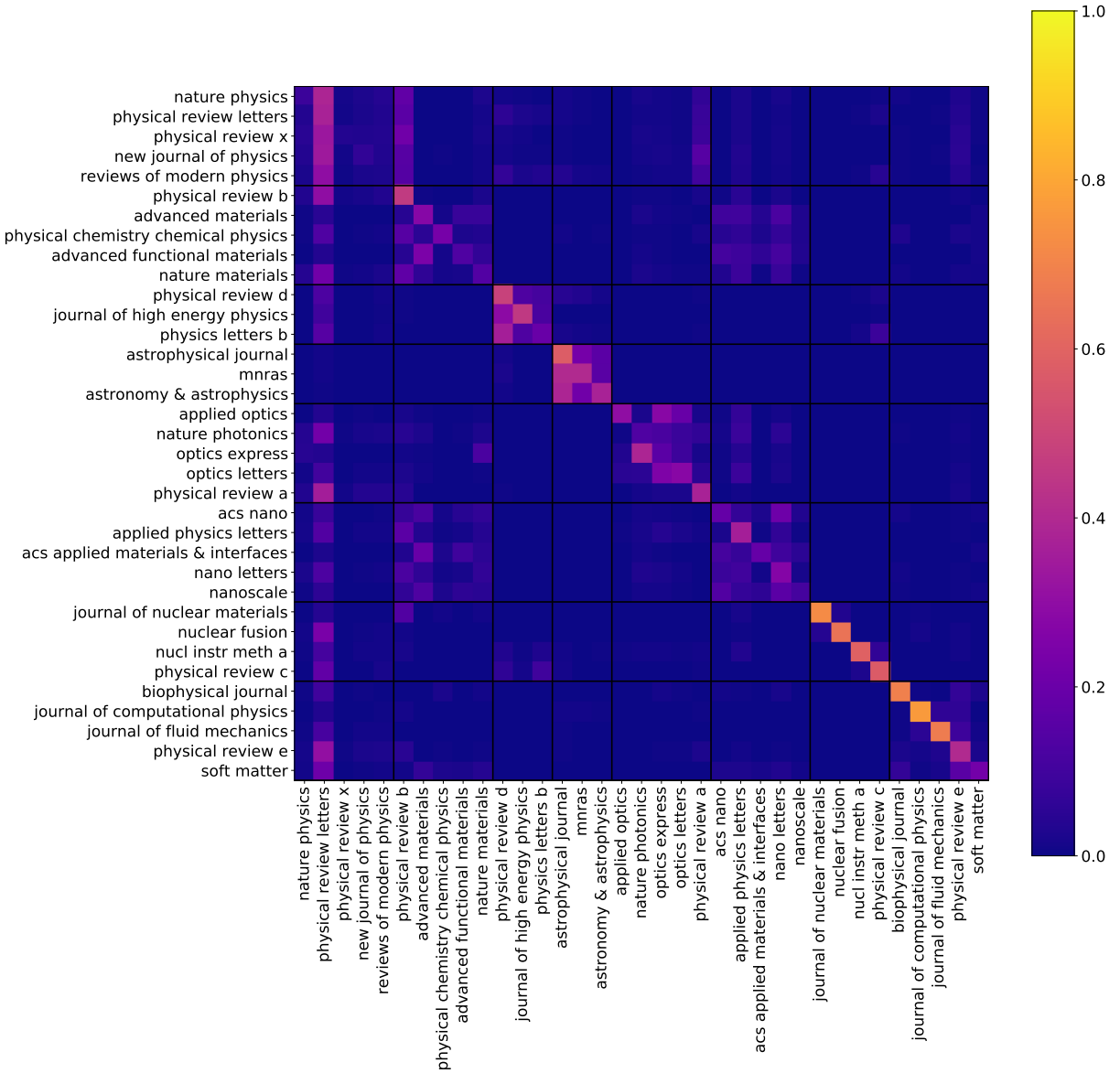

FIG. S5. **Final subfield boundaries.** Each matrix element J_{ij} is colored according to the fraction of citations given by papers in citer journal i to papers in cited journal j, as indicated by the color bar. Subfields are identical to those defined in Fig. S3, with two exceptions: (i) the high energy physics and the astronomy and astrophysics journals have been split into separate subfields, and (ii) *Physics Letters B* has been incorporated into the high energy physics subfield, and removed from the nuclear physics subfield.

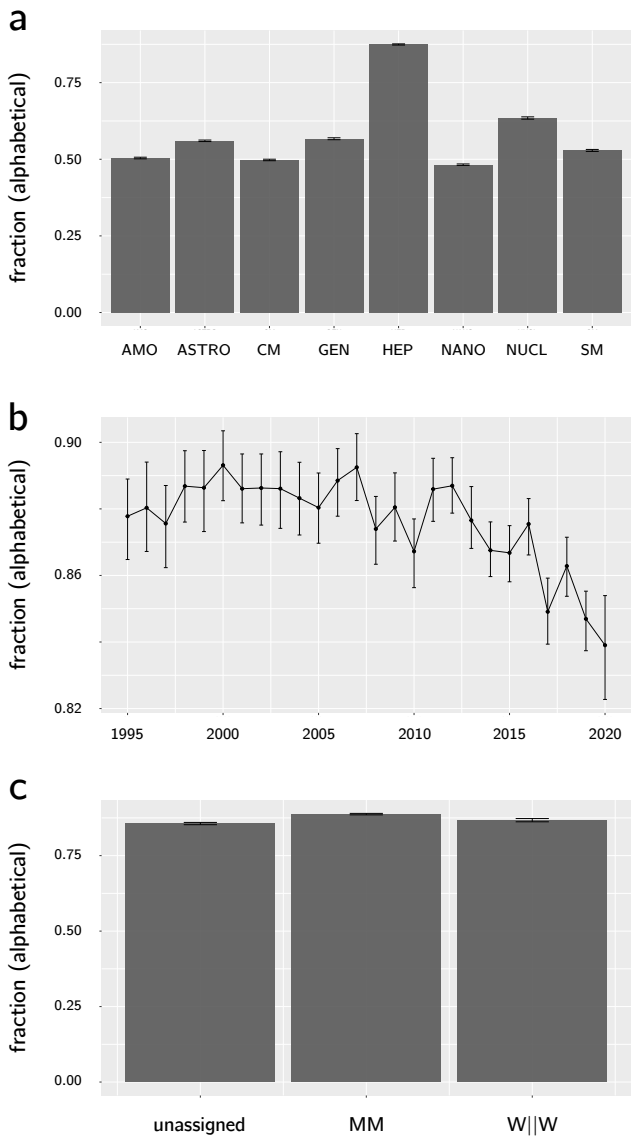

FIG. S6. **High energy physics papers contain alphabetical author lists.** (a) Fraction of all multiple-author papers in each subfield whose first and last authors are in alphabetical order. (b) Fraction of all multiple-author papers in high energy physics whose first and last authors are in alphabetical order, grouped by publication year. (c) Fraction of all multiple-author papers in high energy physics whose first and last authors are in alphabetical order, grouped by author gender category. Papers of unassigned author gender category are shown for completeness. In all panels, error bars represent the 95% CI of each fraction and are calculated from 500 bootstrap resampling iterations.

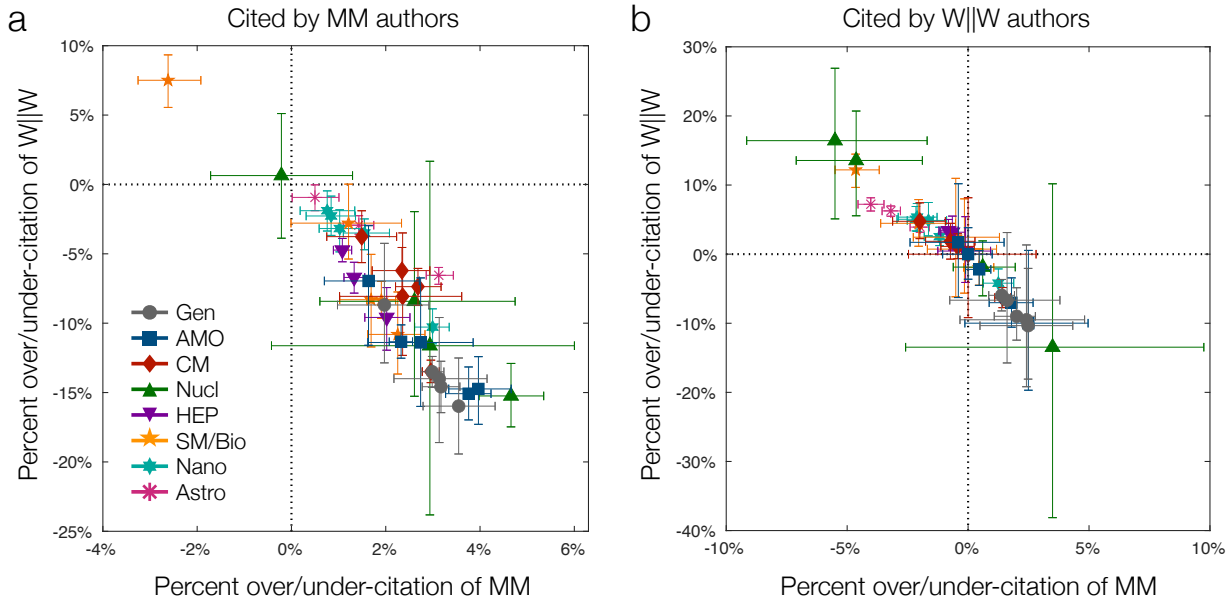

FIG. S7. **Citation behavior grouped according to citer and journal of publication, aggregated over time.** (a) Citation behavior of MM papers published in all journals. (b) Citation behavior of W||W papers published in all journals. In both panels, each journal is a marker, and markers are colored according to journal subfield. Error bars representing the 95% CI of each over-/under-citation value were found via 500 bootstrap resampling iterations. Note that MM papers are far more grouped in the lower right quadrant across the entire set of journals, while W||W papers are more evenly split between the upper left and lower right quadrants. These data show, at the journal level, that authors of MM papers tend to exhibit citation bias that is more in favor of other MM papers, and less in favor of W||W papers, than authors of W||W papers.

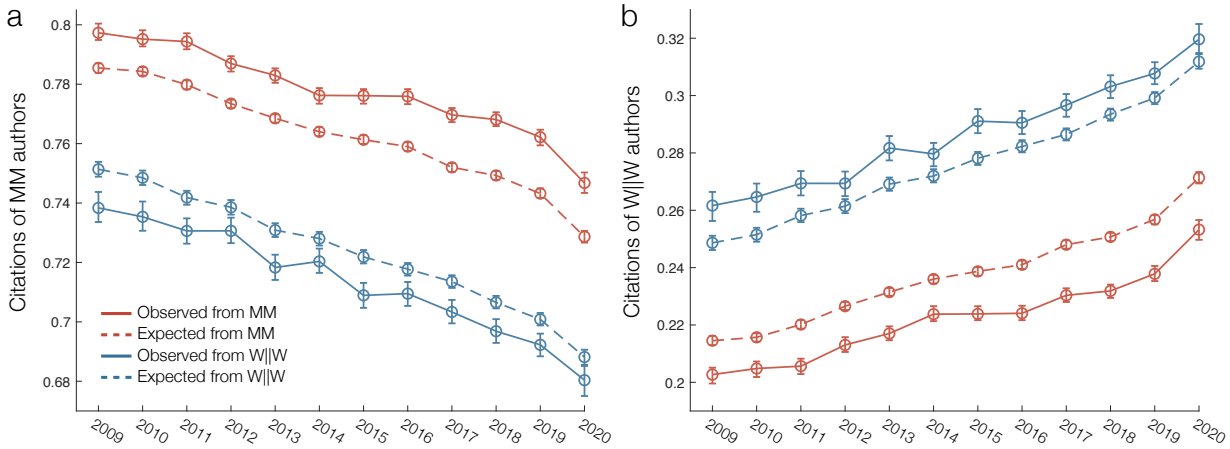

FIG. S8. **Observed and expected citation behavior over time grouped according to citer.** (a) Observed and expected citation proportions given to MM papers over time. (b) Observed and expected citation proportions given to W||W papers over time. Each panel contains two pairs of trends, colored according to the author gender category of the citing team: MM (red) or W||W (blue). Each pair of trends shows the observed yearly proportion of citations given as a fraction of the total number given to either author gender category (solid line), and the equivalent expected yearly proportion according to our paper characteristics model (dotted line). Error bars represent the 95% CI of each proportion, calculated from 500 bootstrap resampling iterations.

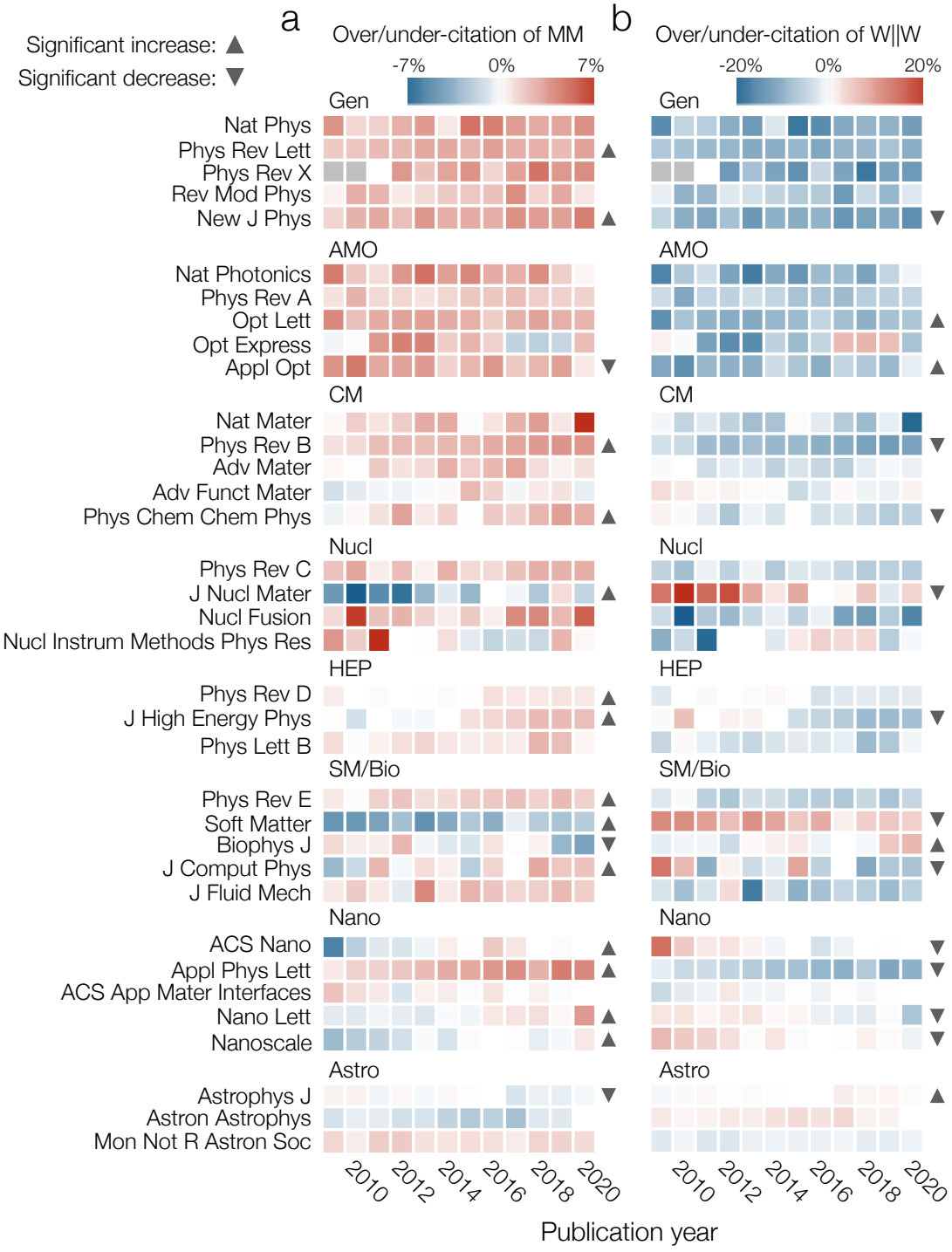

FIG. S9. **Citation behavior over time of individual journals.** (a) Over/under-citation of MM papers by all citing papers in each journal over time. (b) Over/under-citation of W||W papers by all citing papers in each journal over time. In both panels, the reported over-/under-citation values utilize the reference lists of all relevant citing papers, including those of unknown author gender, to increase statistical power. Gray squares occur when journals are not published in any particular year. Up or down arrows next to each journal's citation behavior time series indicate if the relevant over/under-citation trend represents a significant increase or decrease over time, with "significance" determined by a p-value less than 0.05 for a linear fit to each time series.

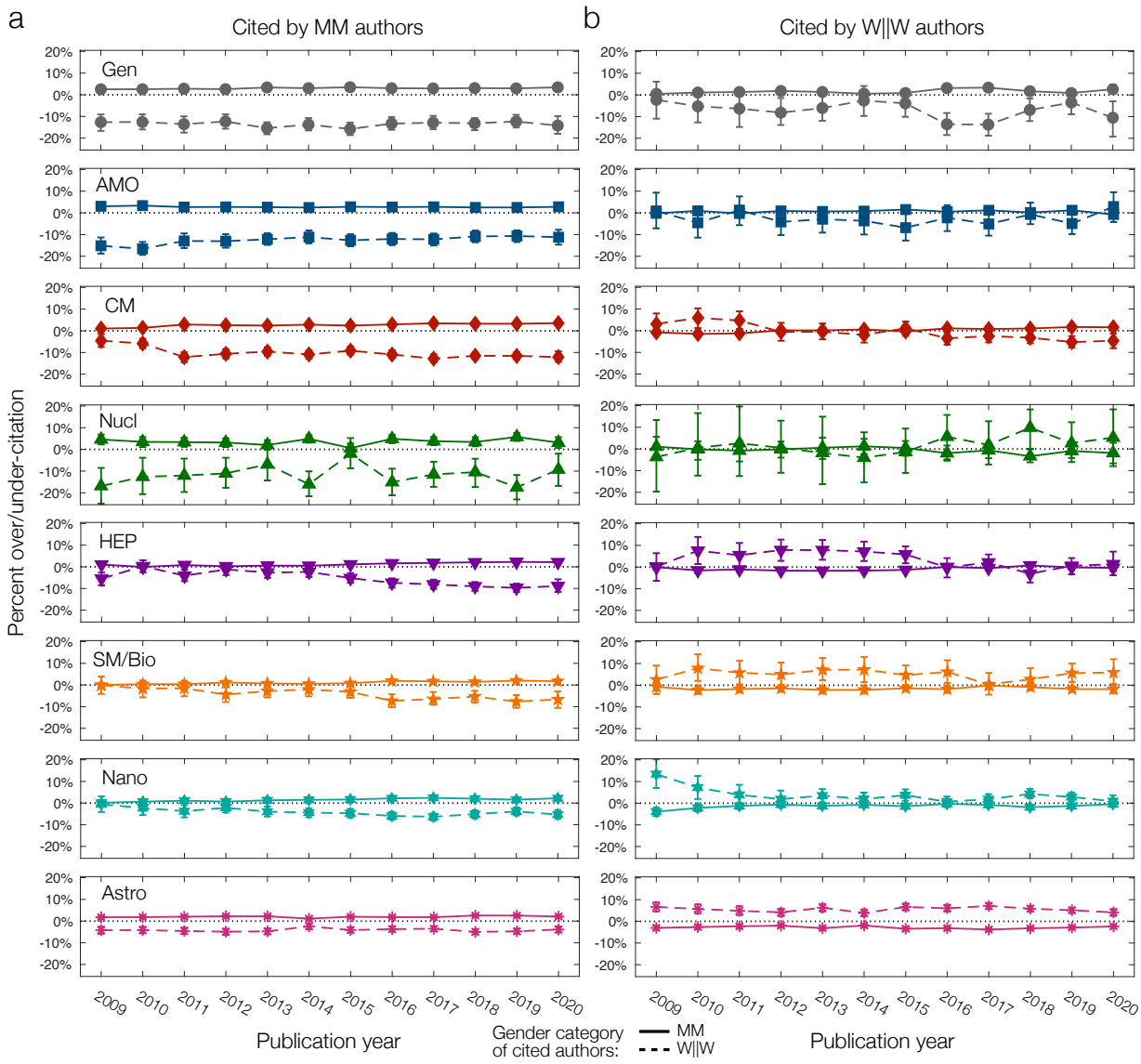

FIG. S10. **Citation behavior over time grouped according to citer and subfield.** (a) Citation behavior over time of MM citer papers in each subfield. (b) Citation behavior over time of W||W citer papers in each subfield. In all panels, solid lines indicate time-varying over/under-citation of MM papers, and dashed lines indicate time-varying over/under-citation of W||W papers. Error bars for each marker represent the 95% CI of each over-/under-citation calculation and were found via 500 bootstrap resampling iterations. Note that MM citers and W||W citers within the same subfield often show markedly different citation behavior over time, with W||W citers usually displaying more equitable citation behavior.

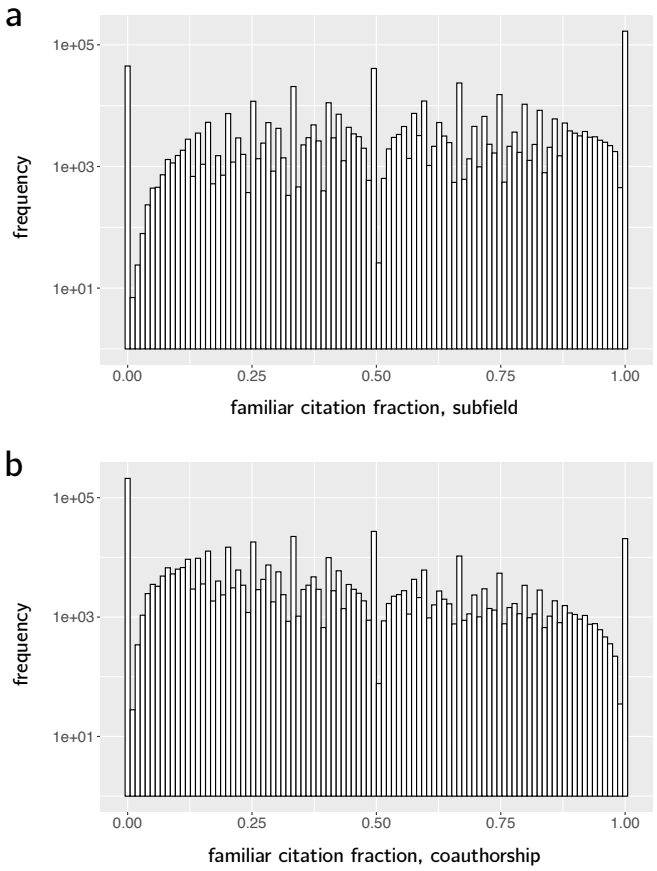

FIG. S11. **Fraction of reference lists defined as "familiar".** (a) Histogram of the fraction of each reference list defined as "familiar" according to whether each cited paper belongs to the same subfield as the citing paper. (b) Histogram of the fraction of each reference list defined as "familiar" according to whether the cited papers share at least one co-author with the citing paper. In both panels, all papers published between 2009 and 2020 with cleaned reference lists of length > 0 are represented.

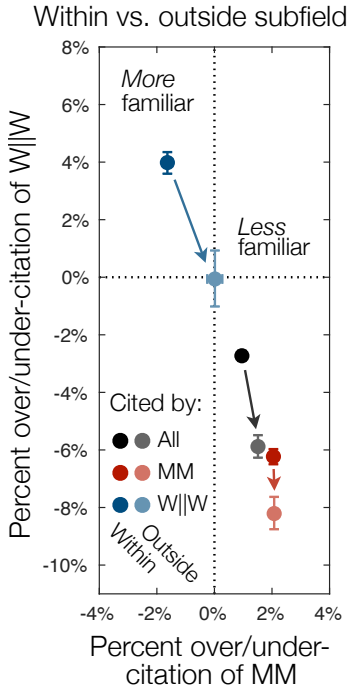

FIG. S12. **Citation behavior varies according to subfield proxy for familiarity with cited work, with subfields defined alternately via citation network clustering.** The familiarity proxy is defined as whether the publishing journals of cited and citing papers fall within the same subfield, with subfields defined according to the citation network clustering shown in Fig. S4. Markers show the citation behavior of (black) all papers written between 2009 and 2020, including those of unknown author gender, (red) the subset identified as MM-authored, and (blue) the subset identified as W||W-authored. Arrows point from over-/under-citation when considering only "familiar" citations to over-/under-citation when considering only "unfamiliar" citations. Error bars representing the 95% CI of each over/under-citation calculation were found via 500 bootstrap resampling iterations.

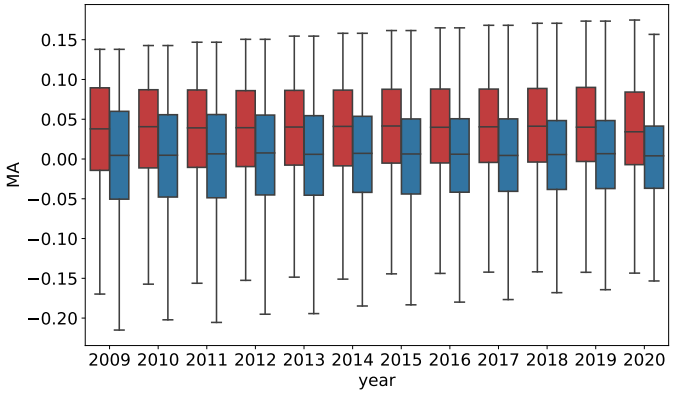

FIG. S13. **Homophilic enhancement exists in co-authorship networks of MM and W||W papers over time.** Box plots show man-author overrepresentation (MA) in the co-authorship networks surrounding MM (red) and W||W (blue) papers aggregated over all years between 2009 and 2020. These box plots remain fairly stable over time, with higher values of MA generally in co-authorship networks surrounding MM papers, and lower values of MA generally in co-authorship networks surrounding W||W papers. The effect is one of homophilic enhancement in each set of co-authorship networks. In all panels, outliers are not shown. All non-review papers published between 2009 and 2020 with cleaned reference lists of length > 0 are represented here. See the main text *Methods* for details regarding calculation of man-author overrepresentation.

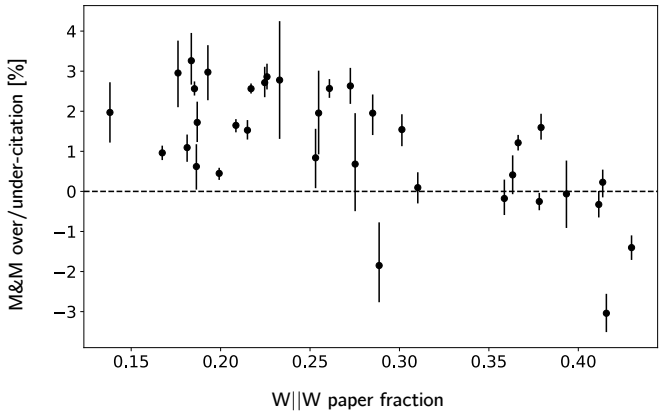

FIG. S14. **Journals with higher proportions of W||W papers published between 2009 and 2020 generally exhibit lower MM over/under-citation.** Each marker represents a journal; error bars representing the 95% CI of each over/under-citation value were found via 500 bootstrap resampling iterations. Proportions of W||W papers are reported with respect to total MM and W||W papers in each journal, and only those with cleaned reference lists of length > 0 published between 2009 and 2020 are represented here.

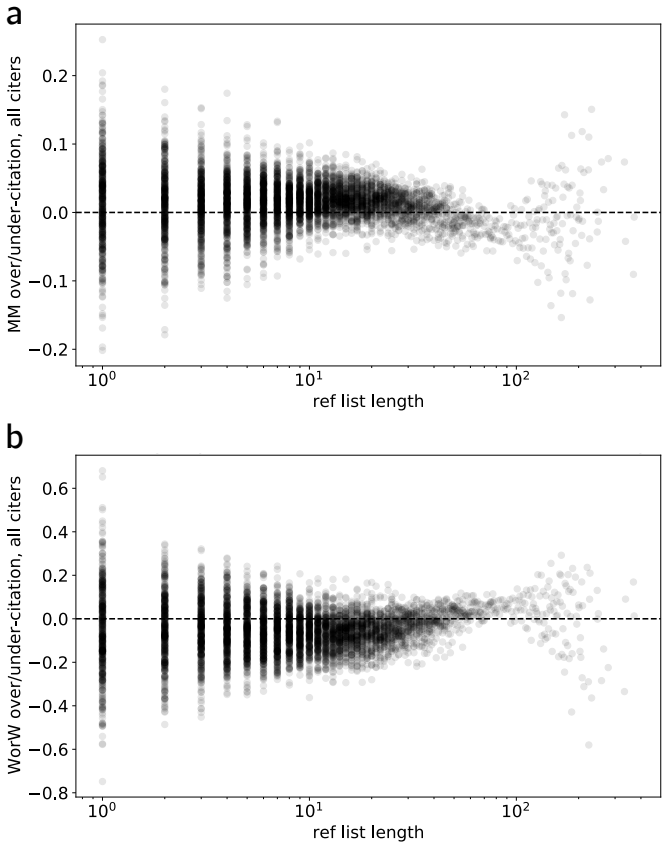

FIG. S15. **Citation bias is correlated with reference list length.** (a) Papers with longer reference lists tend to exhibit lower MM over/under-citation. (b) Papers with longer reference lists tend to exhibit higher W||W over/under-citation. Each data point shows citation bias aggregated over paper subsets of maximum size 100 at each reference list length. All non-review papers published between 2009 and 2020 with cleaned reference lists of length > 0 are represented here.

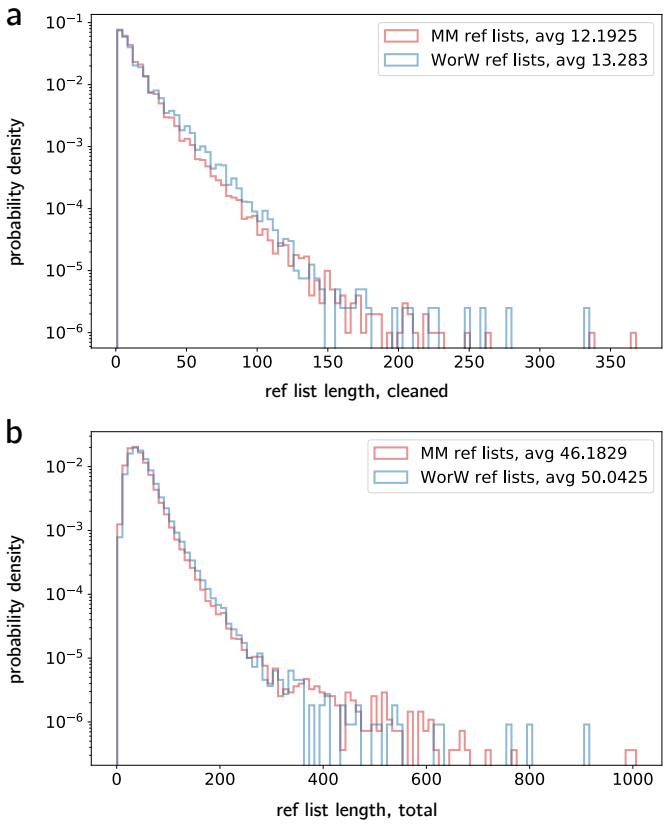

FIG. S16. **W||W citing papers contain longer reference lists.** Each panel shows probability density distributions of reference list length for MM non-review papers (red) and $W||W$ non-review papers (blue) published between 2009 and 2020 with "cleaned" reference lists of length > 0. Cleaned reference lists, as noted in the main text and used by default for all other analyses, consist only of references that are (i) published in the journals and year-range considered in this paper, (ii) not self-citations, and (iii) whose author gender category is identified. Panel (a) shows lengths of cleaned reference lists of all papers, and panel (b) shows lengths of total (non-"cleaned") reference lists of all papers, to demonstrate that the longer reference list length effect for $W||W$ papers is not due to the cleaning process. Average reference list length values are indicated in each panel legend, showing that cleaned $W||W$ reference lists contain about one more reference, on average, than cleaned MM reference lists; and total $W||W$ reference lists contain about four more references, on average, than total MM reference lists.

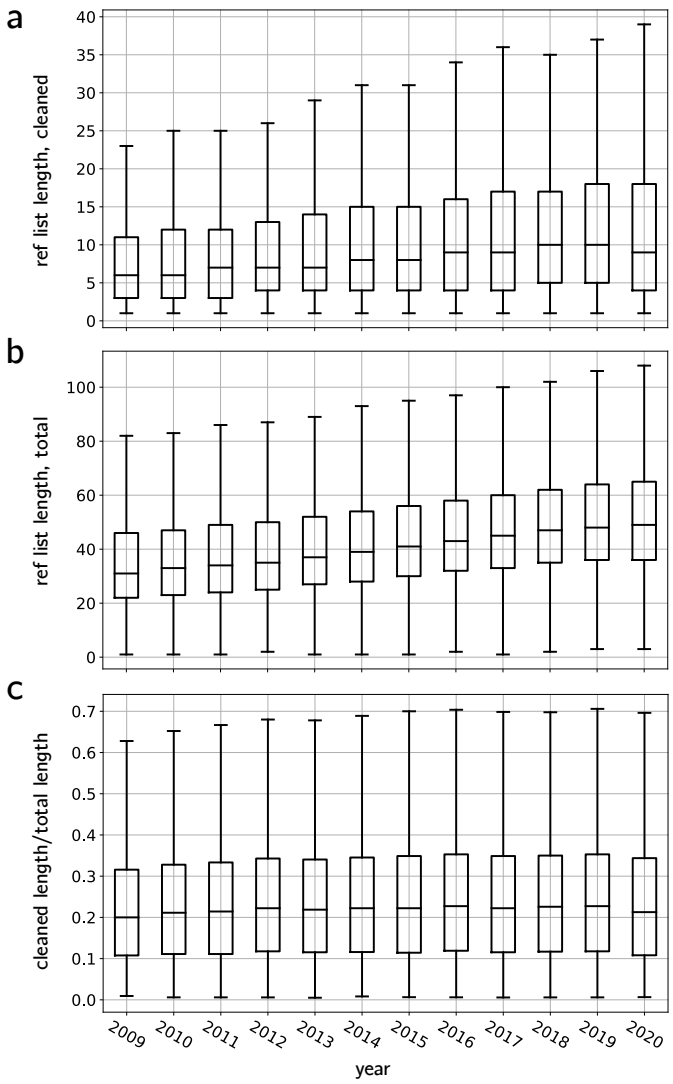

FIG. S17. **Reference list length increases over time.** (a) Box plots of "cleaned" reference list length aggregated over all years between 2009 and 2020. (b) Box plots of total (non-"cleaned") reference list length aggregated over all years between 2009 and 2020. (c) Box plots of the fraction of each total reference list that remains in its cleaned version, aggregated over all years between 2009 and 2020. These box plots remain fairly stable over time, indicating that the cleaning process does not have egregious time-varying effects on reference lists. In all panels, outliers are not shown. All non-review papers published between 2009 and 2020 with cleaned reference lists of length > 0 are represented here.

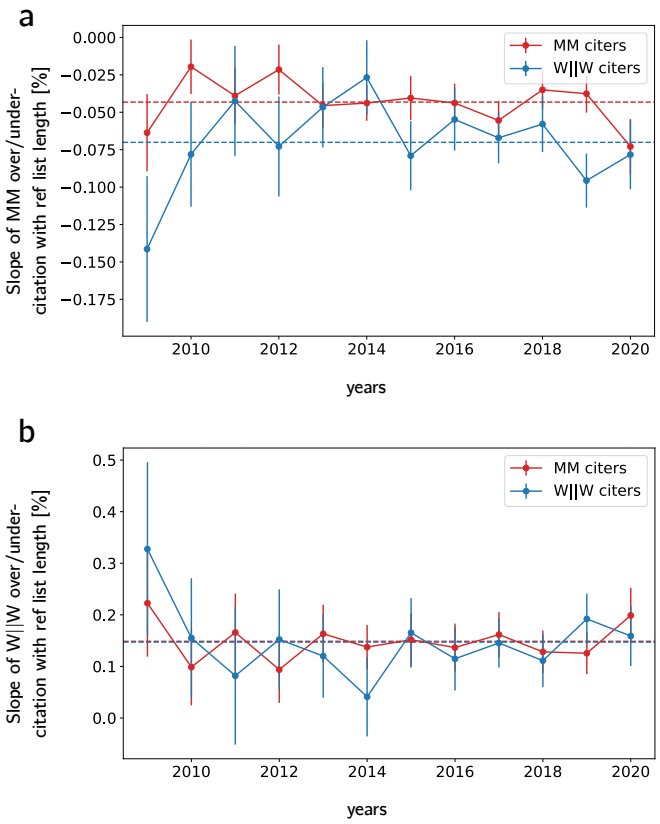

FIG. S18. **Slope of linear fits of over/under-citation with reference list length.** (a) Slopes of lines fit via ordinary least-squares to MM over/under-citation as a function of reference list length. (b) Slopes of lines fit via ordinary least-squares to W||W over/under-citation as a function of reference list length. Each fit considers only citing papers within a single publishing year that are authored by MM teams (red) or W||W teams (blue). Fits are performed on data points that indicate aggregated citation bias across reference lists at each reference list length. Aggregation takes place over paper subsets of maximum size 10 at each reference list length, and only papers with reference lists containing 20% or more within-database and author-gender-categorized citations are fit. Error bars represent the 95% CI on the fitted slopes. Dotted lines are averages taken across year. Consistent negative slopes of MM over/under-citation with reference list length over time, and consistent positive slopes of W||W over/under-citation with reference list length over time, indicate that longer reference lists display less citation bias.

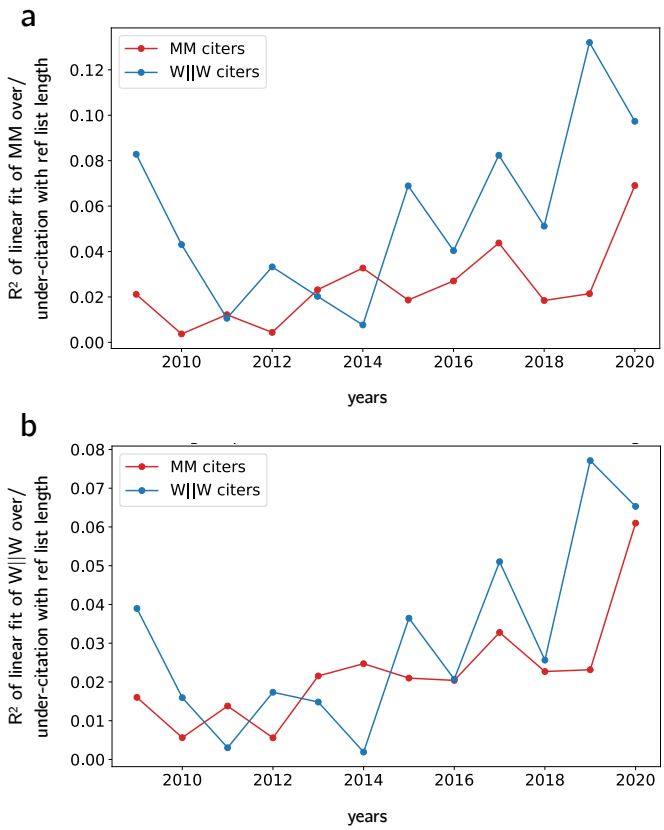

FIG. S19. **Coefficient of determination of linear fits of over/under-citation with reference list length.** (a) Coefficient of determination, R^2, of lines fit via ordinary least-squares to MM over/under-citation as a function of reference list length. (b) Coefficient of determination, R^2, of lines fit via ordinary least-squares to W||W over/under-citation as a function of reference list length. Each fit considers only citing papers within a single publishing year that are authored by MM teams (red) or W||W teams (blue). Fits are performed on data points that indicate aggregated citation bias across reference lists at each reference list length. Aggregation takes place over paper subsets of maximum size 10 at each reference list length, and only papers with reference lists containing 20% or more within-database and author-gender-categorized citations are fit.

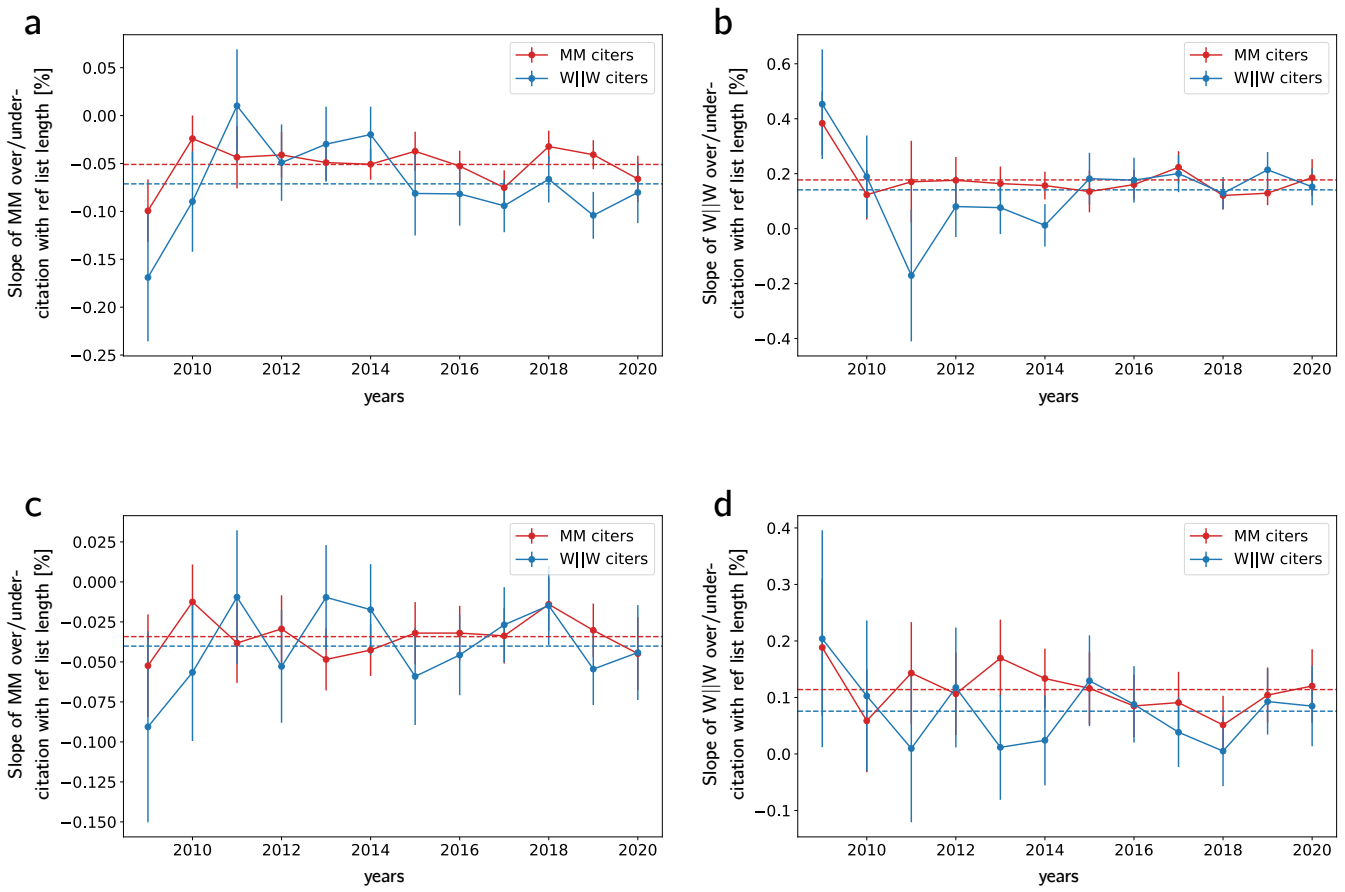

FIG. S20. **Slope of linear fits of over/under-citation with reference list length: Alternate analyses.** (a-b) Slopes of lines fit via ordinary least-squares to (a) MM over/under-citation and (b) W||W over/under-citation as a function of reference list length, when only papers with reference lists containing 40% or more within-database and author-gender-categorized citations are fit. (c-d) Slopes of lines fit via ordinary least-squares to (c) MM over/under-citation and (d) W||W over/under-citation as a function of reference list length, when all relevant papers are fit. Each fit considers only those citing papers within a single publishing year that are authored by MM teams (red) or W||W teams (blue). Fits are performed on data points that indicate aggregated citation bias across reference lists at each reference list length. Aggregation takes place over paper subsets of maximum size 10 at each reference list length. Error bars represent the 95% CI on the fitted slopes. Dotted lines are averages taken across year. Trends are qualitatively consistent across panels (a) and (c), and panels (b) and (d), indicating the stability of this analysis with respect to the paper filtering procedure used.

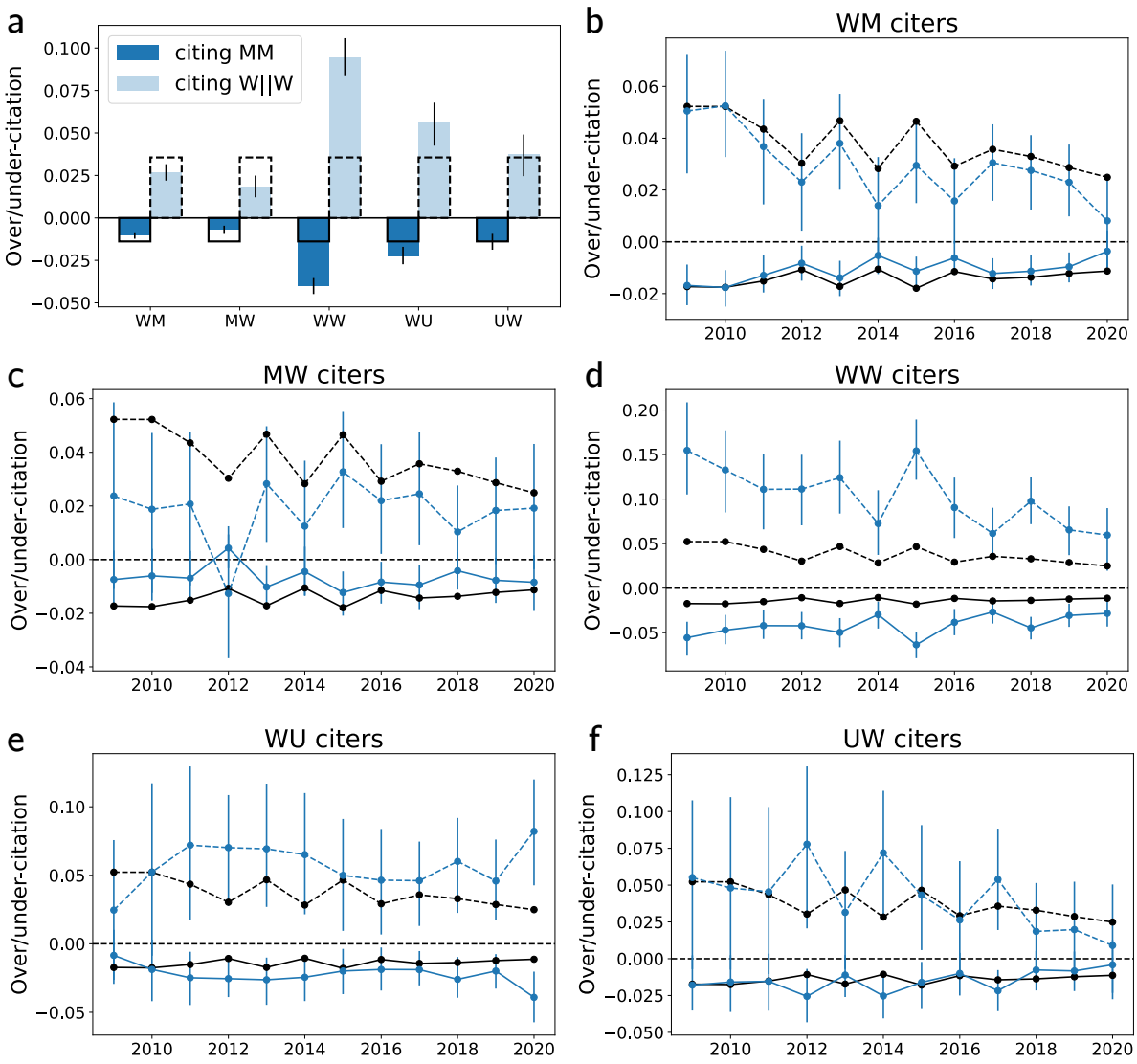

FIG. S21. **Citation behavior of sub-categories of W||W papers.** (a) Over-/under-citation behavior of all sub-categories of W||W papers, calculated separately for papers written by each sub-category published between 2009 and 2020. For every sub-category, the over-/under-citation behavior of all papers in the W||W category is shown for comparison, in black solid and dotted lines (over-/under-citation of MM papers and W||W papers, respectively). (b-f) Over-/under-citation behavior of all sub-categories of W||W papers over time. Citer sub-categories are specified at the top of each plot. Solid lines indicate over-/under-citation of MM papers, and dotted lines indicate over-/under-citation of W||W papers. For every sub-category, the over-/under-citation behavior of all papers in the W||W category is shown for comparison, in black solid and dotted lines (over-/under-citation of MM papers and W||W papers, respectively). In all panels, error bars representing the 95% CI of each over-/under-citation calculation were computed via 500 bootstrap resampling iterations. Error bars for the citation behavior of the larger W||W paper category are not shown for clarity.

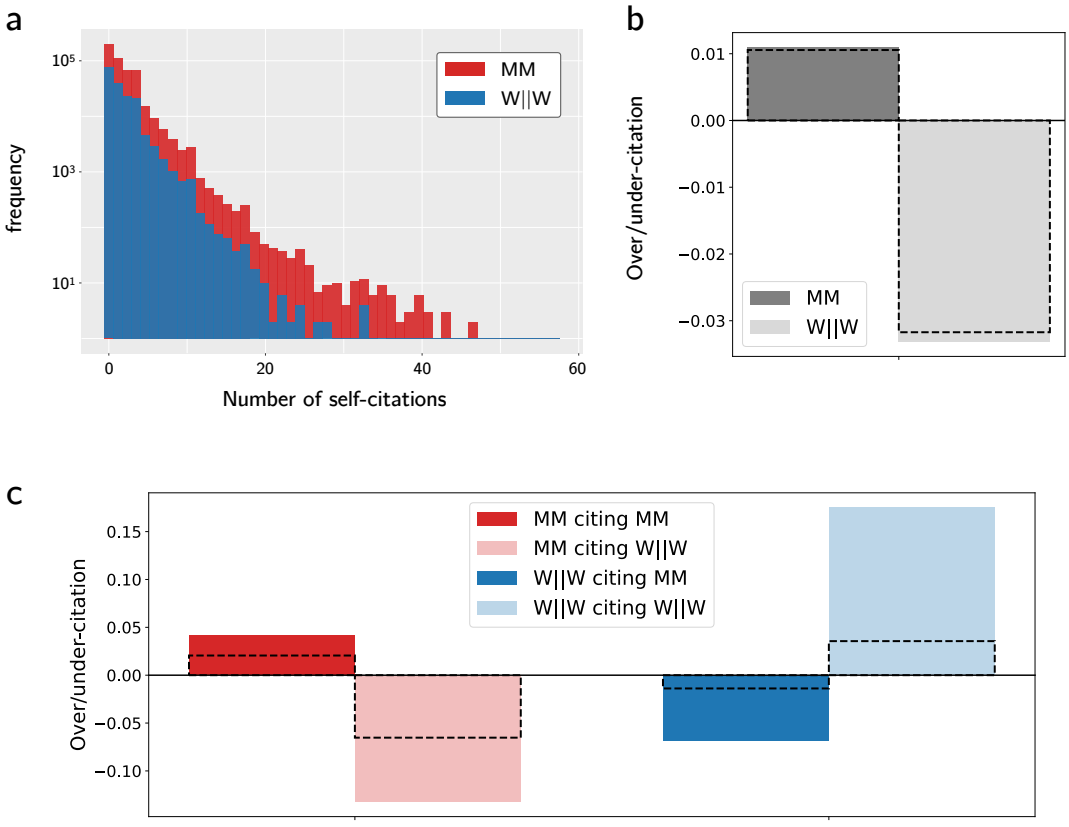

FIG. S22. **Effects of self-citation on citation behavior.** (a) Overlaid histograms of the number of self-citations in all MM and W||W papers in the data set. Note that the y-axis begins at 1, and thus bins of frequency 1 are not shown. (b) Over-/under-citation of MM and W||W papers excluding self-citations (dotted) and including self-citations (colored), calculated for all citing papers published between 2009 and 2020. (c) Over-/under-citation of MM and W||W papers excluding self-citations (dotted) and including self-citations (colored), calculated for MM citing papers and W||W citing papers published between 2009 and 2020. In all panels, error bars representing the 95% CI of each over-/under-citation calculation were computed via 500 bootstrap resampling iterations. Error bars for the citation behavior of the larger W||W paper category are not shown for clarity.

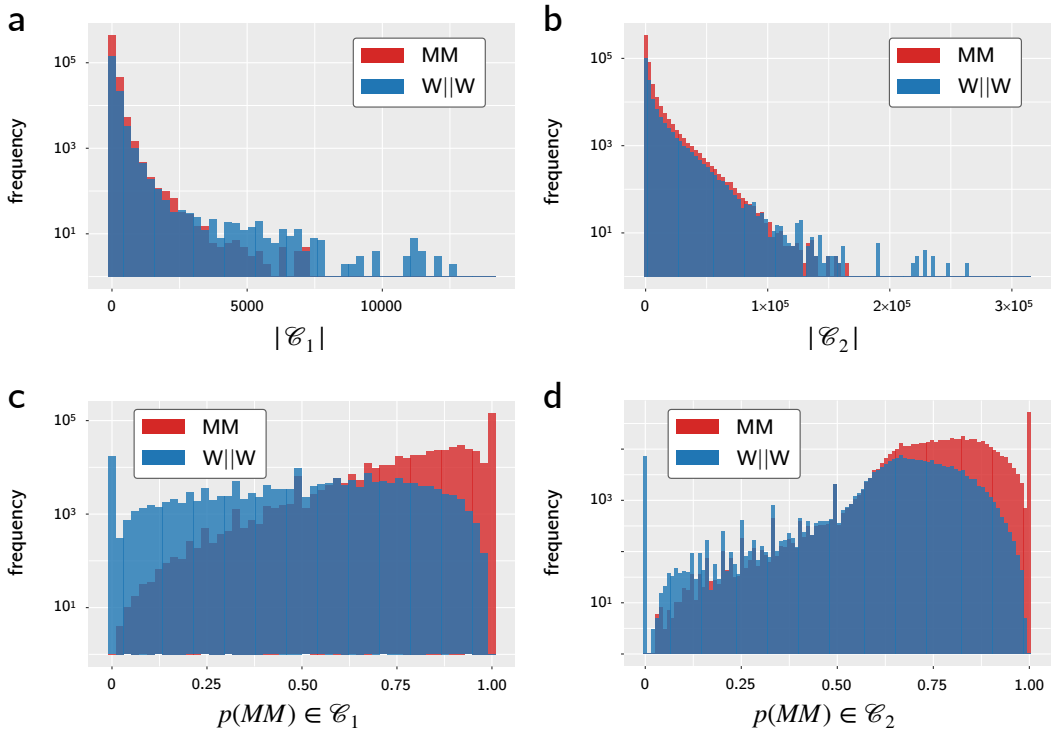

FIG. S23. **Co-authorship neighborhood size and makeup varies according to author gender category.** (a) Overlaid histograms of neighborhood size $|\mathcal{C}_1|$ for neighborhoods surrounding all MM and W||W papers in the data set. (b) Overlaid histograms of neighborhood size $|\mathcal{C}_2|$ for neighborhoods surrounding all MM and W||W papers in the data set. (c) Overlaid histograms of the fraction of MM papers in the neighborhood $|\mathcal{C}_1|$ surrounding all MM and W||W papers in the data set. (d) Overlaid histograms of the fraction of MM papers in the neighborhood $|\mathcal{C}_2|$ surrounding all MM and W||W papers in the data set. Definitions of $\mathcal{C}_1$ and $\mathcal{C}_2$ are provided in the text, Section S3E. Note that in all panels the y-axis begins at 1, and thus bins of frequency 1 are not shown.

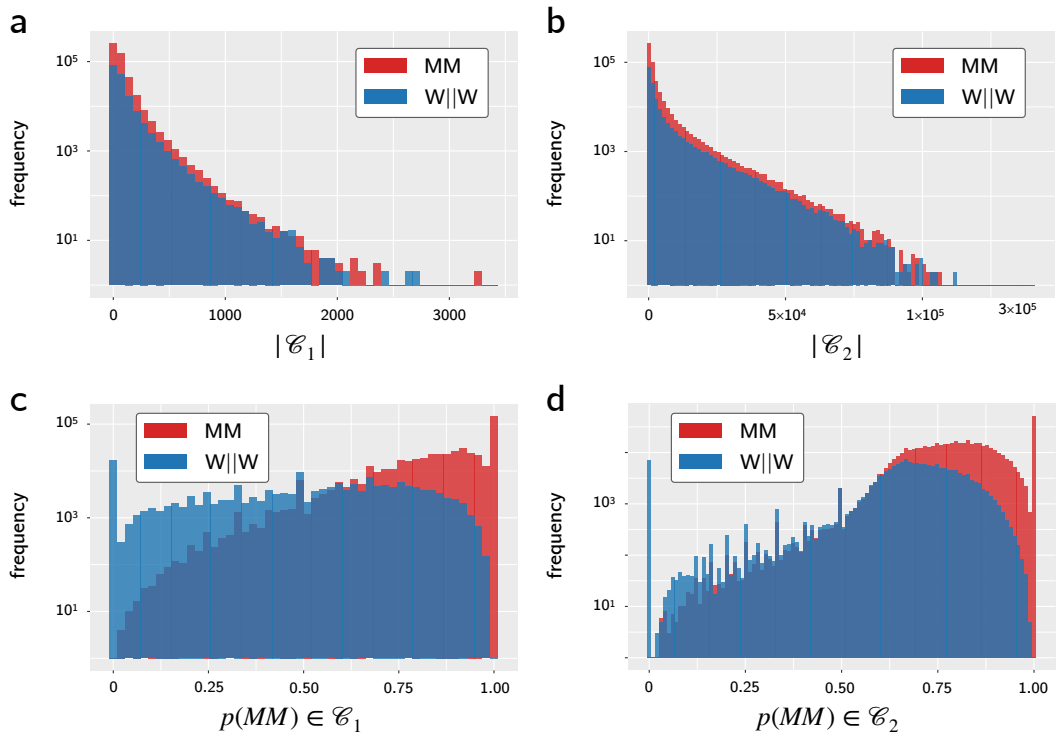

FIG. S24. **Co-authorship neighborhood size and makeup varies according to author gender category, even for papers with a limited number of co-authors.** (a) Overlaid histograms of neighborhood size $|\mathcal{C}_1|$ for neighborhoods surrounding all MM and W||W papers in the data set with author lists of length < 20. (b) Overlaid histograms of neighborhood size $|\mathcal{C}_2|$ for neighborhoods surrounding all MM and W||W papers in the data set with author lists of length < 20. (c) Overlaid histograms of the fraction of MM papers in the neighborhood $|\mathcal{C}_1|$ surrounding all MM and W||W papers in the data set with author lists of length < 20. (d) Overlaid histograms of the fraction of MM papers in the neighborhood $|\mathcal{C}_2|$ surrounding all MM and W||W papers in the data set with author lists of length < 20. Definitions of $\mathcal{C}_1$ and $\mathcal{C}_2$ are provided in the text, Section S3E. Note that in all panels the y-axis begins at 1, and thus bins of frequency 1 are not shown.