Gender Homophily: In-Group Citation Preferences and the Gender Disadvantage

Sifan Zhou[a]*, Sen Chai[b], Richard B. Freeman[c]

Sep., 2023

[a]Department of Public Finance, School of Economics, and the Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, 361005, China.

[b]Strategy and Organisation, Desautels Faculty of Management, McGill University, Montreal, QC, Canada.

[c]Department of Economics, Harvard University and NBER, Cambridge, MA 02138, USA.

*Corresponding author. Email: zhou_sifan@xmu.edu.cn. Phone number: (+86)159-6037-5221

Keywords

knowledge flow, forward citation, gender inequality, gender homophily, field & cohort heterogeneity

JEL Classifications Codes

J16, O32, O33

Acknowledgments & Funding Sources

The authors thank Matt Marx, Michaël Bikard, Griffin Weber, and Raviv Murciano-Goroff for helpful suggestions and comments, as well as participants of the 2022 Wharton Technology & Innovation Conference, the 2022 Strategic Management Society Annual Meeting, and the McGill Desautels Organizational Behaviour Working Group Meeting.

This project was supported by funding from the National Natural Science Foundation of China project grant 71804155, the CERESSEC (Le Centre de Recherche de l'ESSEC) project grant 043-288-2-1-06-P-1, and the China Fundamental Research Funds for the Central Universities grant No. 2072018104

Gender Homophily: In-Group Citation Preferences and the Gender Disadvantage

Abstract

Based on an extensive sample of articles in the life sciences, we find that gender homophily in forward citations is substantial: compared to men-led articles (i.e., those with men as either the first or last author), women-led articles receive fewer forward citations from subsequent men-led articles and more forward citations from subsequent women-led articles. This occurs across life science fields with varying gender ratios. Forward citations flow differentially to papers led by women versus men for a variety of reasons, including the detailed field and scientific concepts covered in the articles, the journals in which they are published, article length, authors' research experience, and the size of the author team. After accounting for this extensive set of factors, we find some forward citations appear to be driven by gender citation homophily – that is, gender alignment between citing and cited authors. This pattern greatly disadvantages women in fields where they are underrepresented, leading to a gender citation gap, compared to more gender-balanced fields, where the gap is shrinking. We also find that articles written by more recent cohorts of scientists are subject to less gender citation homophily than earlier cohorts. Investigation into potential pathways by which gender citation homophily operates suggests it stems from gendered specialization in research niches and, to a lesser extent, from gender homophily in professional connections among scientists, as opposed to from direct discrimination against unknown authors based on gender inferred from their names. Since gender homophily in citations impedes gender-indifferent knowledge flow in most fields, its adverse impact on science likely includes not only slowing women's careers but also creating a less efficient diffusion of knowledge and recombination of work from earlier papers into newer work.

1. Introduction

While the proportion of female students with degrees in science has increased greatly in recent decades, women's careers have advanced more slowly, due in part to women publishing fewer articles than men at key career points (Ceci et al., 2014; Cole and Zuckerman, 1984; Lerchenmueller and Sorenson, 2018; Xie and Shauman, 1998) and to women gaining fewer article citations than men (Beaudry and Larivière, 2016; Caplar et al., 2017; Chatterjee and Werner, 2021; Huang et al., 2020; Larivière et al., 2013; Long, 1992; Maliniak et al., 2013). High-profile cases in science history (Isaacson, 2021; Klug, 1968) and large-sample studies (Ross et al., 2022) further show that women are less likely than men to be credited with authorship for their scientific contributions. Given that citations are the primary indicator of scientific prowess, lower recognition via citations and credit for work[1] contributes to high career attrition and the under-representation of women in senior positions (Ceci and Williams, 2011; Huang et al., 2020; Larivière et al., 2013).

This paper examines differences in the *forward citations* (references made in subsequent articles to a cited article) received by papers written by men and women in the life sciences, a broad scientific area where women now earn the majority of new PhDs, and assesses the contribution of gender homophily in forward citations to the male-female citation gap. Given the greater number of male scientists and publications by male authors in most fields, gender homophily in citations contributes to the male-female citation gap for articles that are otherwise observationally similar. This slows female progress in science and, to the extent that scientists build disproportionately on the work of researchers of their own gender rather than that of a more diverse research community, distorts and slows the advance of science as well.

Our analysis builds on earlier work (Dion et al., 2018; Dworkin et al., 2020; Ghiasi et al., 2018; Hutson, 2002; Teich et al., 2022; Wang et al., 2021) documenting *backward citations* (actual reference lists of articles) written by men consist of a higher percentage of articles written by other men compared to *predicted* reference lists constructed based on topic relevance.[2] In-group preferences for referencing articles by authors of one's own gender contribute to differences in forward citations, but the link between gender homophily in backward references to gender effects in forward citations is indirect and underexplored.

---

[1] Going beyond citations, innovations by women are less disseminated (Vásárhelyi et al., 2021), more undervalued (Hofstra et al., 2020), and less likely to be turned into applicable technologies (Bikard and Fernandez-Mateo, 2022).

[2] To predict the gender composition of reference lists in the absence of gender homophily as a benchmark for comparison, some articles accounted for five characteristics of the cited articles, including publication year, author number, author seniority, journal, and article type (Dworkin et al., 2020; Teich et al., 2022; Wang et al., 2021), while others considered nearest neighbor articles based on the similarity of the title and abstract to account for the topic (Ghiasi et al., 2018).

The effect of backward citation gender homophily on the forward citations received by female-written compared to male-written papers depends on the magnitude of women and men's homophily in citations, the gender distribution in different scientific communities, and the interaction between the two. For instance, in a field where few women author papers and thus make references, even mild gender homophily in citations could yield a nontrivial female disadvantage in forward citations, as the extra citations that male-authored papers receive from subsequent male-led articles will dwarf the extra citations that female-authored papers receive from subsequent female-led articles. Meanwhile, in a field where both genders exhibit strong homophily but the gender composition is more balanced, their homophily could offset each other and thus result in a small gender gap, if any, in overall forward citations. Given the importance of citations in academic evaluations, examining how gender citation homophily affects the gender gap in forward citations is critical to pinning down this impediment to women's continuing progress in their academic careers.

Our paper focuses on research articles in the life science (defined to include biological, agricultural, environmental, and health sciences), for three reasons. First, life science is the largest major area to be populated by female scientists. Second, it has a sizeable number of disciplines, which allows us to compare gender homophily among fields with differing proportions of women authors, something that earlier work on gender differences in individual fields could not do (e.g. Hutson, 2002; Maliniak et al., 2013 vs. Montpetit et al., 2008). Third, it exemplifies the conundrum between the progression of women obtaining the majority of PhDs and their continued minority status in academic positions.

Given the importance of first and last authors in scientific work, we classify articles by the gender of its first author, and by that of its last author.[3] Since papers and authors differ in many characteristics beyond gender, we estimate the "pure" effect of gender and gender homophily on citations by comparing articles and authors that are as observationally as similar as possible. We find that even among observationally equivalent papers and authors, gender homophily in citations is substantial. Compared to men-led articles, women-led articles receive fewer forward citations from subsequent men-led articles while receiving more forward citations from subsequent women-led articles. This is true across life science fields with varying gender ratios and in our overall sample. This pattern not only greatly disadvantages woman in fields where they are underrepresented, but also yields gender bias in knowledge flow in almost all fields. Across cohorts, we find that younger cohorts of scientists face less gender homophily in receiving forward citations than older ones. Looking into pathways by which gender citation homophily could operate, we find it tends to stem from gendered specialization in niches of research topics and, to a lesser extent, from gender homophily in professional connections among

[3] In robustness tests, we also examine the gender of authors at other positions and obtain results consistent with those obtained from examining the gender of first or last authors of articles.

scientists, rather than from direct discrimination against unknown authors based on the gender inferred from their names.

Given the prevalent use of forward citations to assess the scientific contributions of articles and evaluate scientists in hiring, promotion, and funding decisions (Bhattacharya and Packalen, 2020; Bol et al., 2022; Graddy-Reed et al., 2019; Xie and Shauman, 2003, 1998), this in-group preference will strengthen a minority group's minority status at various stages of academics' careers. To the extent that forward citations are the "paper trail" left to identify knowledge diffusion and professional interactions, our findings suggest a more general phenomenon of in-group preference that hampers gender-indifferent knowledge flow and reinforces gendered choices in research areas (Koning et al., 2021), even when the scientific community appears to be gender-equal. Our evidence from an extensive dataset across life science fields can thus serve as another reference point for policymakers and administrators when evaluating and funding academics as well as when enacting equity, diversity, and inclusion policies (Graddy-Reed and Lanahan, 2023).

2. The Context: More Influx of Women in Science Degrees than in Publications and Citations

While women have historically made up a minority of scientists, they have been catching up to men or surpassing them as graduate students and PhDs in some fields, particularly in the life sciences. Figure 1A shows that in the early 1970s, less than 20% of newly graduated life science PhDs were women; by 2005, this share has surpassed 50% and reached 55.8% by 2020. Moreover, even though women's share of junior faculty positions increased with the female share of doctoral degrees, resulting in a constant gap between degrees and junior faculty positions of about 5 percentage points, the gender gap is higher in the upper parts of the academic career ladder (Ceci et al., 2014; Larivière et al., 2013).

[Insert Figure 1 about here]

The post-PhD gap in women's career progression has been linked to research productivity, commonly measured by some function of the number of papers published and citations to those papers, such as total forward citations over one's academic career or the Hirsch Index (Ellison, 2013; Hirsch, 2005; Van Raan, 2006). These measures directly affect a researcher's key career events, such as hiring, promotion (including tenure decisions), retention, compensation, and funding. Much of the early literature on the gender gap in academic achievements focused on publication counts as an outcome variable to proxy research productivity (Ceci et al., 2014; Cole and Zuckerman, 1984; Levin and Stephan, 1998; Xie and Shauman, 2003). Studies found that women publish less frequently than men in ways that cannot be fully explained by field, academic rank, teaching burden, fertility, and research funding, creating a "productivity puzzle" in outputs (Ceci et al., 2014; Cole and Zuckerman, 1984; Xie and Shauman, 1998).

A second wave of studies examined the frequency with which the research of women and men is cited. Studies based on large samples of publications across fields[4] tend to find that women are cited less per publication than men (Beaudry and Larivière, 2016; Huang et al., 2020; Larivière et al., 2013), which we also observe in our dataset, as shown in Figure 1B. Studies based on smaller samples of specific fields show that in archaeology (Hutson, 2002) and international relations (Maliniak et al., 2013), women have fewer average citations per article than men, but also find a modest female advantage of citations in political science (Montpetit et al., 2008). Since highly cited articles are often sources of breakthrough discoveries and radical innovations (Chai et al., 2020; Fleming, 2001), the citation gap can be viewed as a precursor to the predominance of male researchers in those applications of new research (Koning et al., 2021).

Digging further into the driving factors of the gender citation gap, studies have identified gender homophily as a potential mechanism. Homophily, which has been extensively studied in the social sciences since the mid-20th century (Lazarsfeld and Merton, 1954), encompasses the idea that "ties between similar people are more frequent than between dissimilar people" (Ertug et al., 2018, p. 219). Existing studies have shown that homophily manifests in various social interactions, including in scientific collaborations, and along various dimensions (McPherson et al., 2001), such as ethnicity (Freeman and Huang, 2015), age (Reagans, 2011), and nationality (Gibson and Zellmer-Bruhn, 2001; Joshi et al., 2002). Most studies on gender homophily in citations commonly focus on *backward* citations (Dion et al., 2018; Dworkin et al., 2020; Hutson, 2002; Teich et al., 2022; Wang et al., 2021). They predict the gender composition of reference lists in the absence of gender homophily as a benchmark for comparison and show how the gender composition of actual reference lists differs from the benchmark depending on the gender of the authors who make the references. Most predictions are based on characteristics of cited articles, including publication year, author number, author seniority, journal, and article type (Dworkin et al., 2020; Teich et al., 2022; Wang et al., 2021), while others use nearest neighbor articles based on similarity of nouns in the article's title and abstract to account for topic area (Ghiasi et al., 2018).

Although the existence of gender homophily in backward referencing behavior has been clearly confirmed, its impact on the gender gap in forward citations has only been implied and not directly estimated. In our analysis, we explicitly quantify the link between the gender-homophilic referencing behavior and the gender gap reflected in forward citations. Since the number of forward citations garnered

[4] In Larivière et al. (2013) and Beaudry and Larivière (2016), each article's number of forward citations is divided by the average number of citations received by articles in the same discipline published that year. The gender citation gap is then estimated with the pooled sample of all fields based on these discipline-normalized citations. In Huang et al. (2020), the gender citation gap is estimated for each broad field.

is the ubiquitous measure used in academic evaluations, rather than backward citations made, the way in which forward citations are influenced by factors other than articles' scientific content, such as gender homophily, have significant implications for the gender gap observed in academic careers.

Another feature of the existing literature on gender homophily in citations is that they have typically focused on single fields, such as archaeology (Hutson, 2002), international relations (Maliniak et al., 2013), neurosciences (Dworkin et al., 2020), political science (Dion et al., 2018), and communications (Wang et al., 2021), etc. Due to the research design of these field-specific studies, little is known about how gender homophily and the gender gap in citations it creates vary across fields with different gender ratios and whether it stems from gendered choices in research interests. In this paper, we address these gaps in the literature by exploring the gender gap in research productivity from the perspective of forward citations. We also address whether and how the relative degree of homophily exhibited by the two genders and the resulting gender gap in citations change depending on field-level gender representations.

3. Empirical Strategy

We analyze gender differences in citations in the life sciences in three steps. First, we combine paper and citation data from two major bibliometric databases, PubMed and Microsoft Academic Graph (MAG), to measure the citation links between papers. Second, we impute the gender of authors from an algorithm that relates first names to the author's likely gender. And finally, we develop regression analyses, controlling for characteristics of papers and authors, thereby isolating the effect of author gender on the number and gender composition of forward citations.

3.1 Data and Sample

As a first step, we combine the National Library of Medicine's bibliographic database for life science publications, PubMed, with the Microsoft Academic Graph (MAG) database, which contains detailed information on the field and scientific context of the publications.[5] PubMed provides information on publications, including article title, journal title, publication year, author names, author position, publication language, and publication type, in all life science fields. Since 2002, it has reported full first names whenever available,[6] which we use to impute authors' gender. Because PubMed does not provide all the citing-to-cited article linkages,[7] we use those from MAG to calculate an article's forward citations. To determine whether the "John Smith" listed on two articles is the same individual, we use the name-

[5] We use PubMed API to download the data (https://www.nlm.nih.gov/databases/download/pubmed_medline.html). MAG is now available for bulk download at OpenAlex (https://docs.openalex.org/).

[6] Full first names are recorded as long as they appear at the author position in the article. If the full first names are not reported at the author position but appear somewhere else in the article, PubMed does not record it.

[7] PubMed only provides the citing-to-cited article linkage when both the citing article and the cited article are part of PubMed Central, a subset of PubMed.

disambiguated author identifier provided by MAG (Wang et al., 2019). This identifier allows us to map an author's career history from first publication to compute experience and identify self-citations.

We restrict our sample to journal articles published between 2002 and 2017. We compute three-year forward citations and keep articles in our sample even if they receive *no* forward citation in that time window after publication.[8] We match about 95% of articles in PubMed one-to-one to records in MAG and drop the unmatched records as well as matches that are not one-to-one.[9] We further restrict our sample to English publications (93.9% of PubMed articles) where both the cited and citing articles are in English to avoid cross-language barriers in knowledge diffusion. We also restrict our sample to articles with no more than 17 authors (the 99th percentile of the distribution of the number of authors) to rule out the concern that both the nature of research and the effects of authors' gender composition on forward citations may be different for very large teams. We drop articles with missing data.

Finally, to be able to link our analysis to rising female participation in the life sciences in the United States, we analyze cited articles with U.S.-based first authors (28.8% of PubMed articles) while including citing articles from all country affiliations. We were able to assign both first and last author gender for 87% of U.S.-based articles (see the next section for a detailed gender imputation methodology). When we restrict our sample to U.S.-based articles where we can identify the gender of both first and last authors, our final sample consists of 2,432,806 cited articles. Note that although we apply multiple criteria to restrict the final sample of cited articles for analysis, we use all available matched articles between PubMed and MAG from 1975 to 2020 for variable construction, such as tracking authors' publication history and research experience. Appendix Table A1 documents the reasoning for each step of our data-cleaning process and the corresponding attrition rates.

3.2 Gender Imputation

To infer the gender of authors, we extract first name and country information from author-affiliation level data and clean the name, including running initials, suffixes, prefixes, spaces, hyphens, quotes, non-English letters, and multiple first names, etc. We input each first name-country level observation into the Genderize.io API (Santamaría and Mihaljević, 2018) twice – once with country information and once without. In the absence of country information, Genderize.io provides a "global" frequency of the name,[10] the gender, and its probability. With country information, Genderize.io provides

[8] We exclude reviews, books, conference papers, letters, and editorial materials from our sample. Since our data ends in November 2021, publications from 2018 to 2020 are used to compute three-year forward citations for the 2017 cited articles.
[9] MAG provides article-level linkages to records in PubMed. We take advantage of these linkages for matching.
[10] Frequency indicates the popularity of a given name in the Genderize.io database. It is the number of individuals with that given name with known, validated and, confirmed gender by Genderize.io.

a country-specific frequency,[11] the corresponding gender, and probability. The gender imputed with country information, when available, is more accurate than the gender imputed using the name alone. For example, "Jean" is more likely to be a male name in France but female in the United States. Since the rarer a name is, the less confident we are in the provided gender, gender is only inferred when the frequency of a name in the database meets a threshold of ten. We apply this threshold for both global and country-specific imputations of gender. In cases where the global imputed gender of a first name conflicts with its country-specific gender, we use the latter. Finally, to address the problem that different genders may be imputed for the same individual,[12] we use the modal gender as the single imputed gender for each unique author.

Since more than 90% of journal articles in our sample have more than one author, we follow authorship norms in the life sciences by classifying the gender of multi-authored articles as women-led using the gender of either the *last author* – usually the principal investigator and/or head of laboratory – or the *first author* – commonly the main researcher and/or advisee (Baerlocher et al., 2007). In our sample, 26.2% of articles (637,309 articles) have women as *last* author, 36.5% (888,079 articles) have women as *first* author, and 14.5% have women as both.

3.3 Empirical Specification

To estimate the gender gap and degree of gender homophily in receiving forward citations, we use the following linear regression model:

$$citations_{itjfa}^{G} = \beta_0 + \beta_1 female_i + \Gamma X_i + \sigma_f + \tau_a + \lambda_j + \mu_t + \epsilon_{itjfa} \quad (1)$$

where i denotes the article, t the year of publication, j the journal, f the research field or scientific concept, and a the author's affiliation.

The dependent variable, $citations_{it}^{G}$, is the number of forward citations that article i published in year t receives in the three-year window post publication.[13] The superscript G denotes variations of the *forward citation* variable, which we use in different analyses, including: (1) the *total number* of citations an article garners, (2) the number of *self-citations* to the article from any of its authors, (3) the number of non-self-citations from *male-led* articles, and (4) the number of non-self-citations from *female-led* articles.

[11] The imputed country-specific frequency is always smaller than the global frequency; therefore, there are more chances for the API to return a null result when a first name is country restricted.
[12] This may be because a unique author's name is spelled differently, their first name is missing on a publication, or they have multiple affiliations due to career moves or multiple appointments in different countries.
[13] Three-year citations are a good predictor of both five-year and ten-year citations: the correlation between three-year and five-year citations is 0.977, and that between three-year and ten-year citations is 0.888. We repeat all our analyses using five-year and ten-year citation windows and find robust results.

The key independent variable is the inferred gender of the cited article's author, $female_i$. In our main analysis, we use the following two variables: (1) *last author female* equals 1 if the article's last author is a woman, and (2) *first author female* equals 1 if the article's first author is a woman. In either case, a negative estimated β_1 indicates that women-authored articles receive fewer citations compared to men-authored articles. As noted, given potentially wide differences in the characteristics of articles other than the gender of its authors, we developed an extensive list of factors that impact citations and that can be correlated with the author's gender. This includes: a vector of variables for author and article characteristics, X_i; a series of fixed effects σ_f for the *field* of a paper or the *scientific concepts* it studies; a series of *affiliation* fixed effects τ_a to account for potential sorting by gender into different affiliations and the likely effect of affiliations on citations; a set of journal fixed effects λ_j;[14] and publication year fixed effects μ_t. We estimate the model on our sample of 2,432,806 journal articles and report robust standard errors at the journal level.

3.4 Variables and Summary Statistics

Table 1 contrasts observable factors that could impact forward citations for women and men in our dataset. It identifies substantial and statistically significant differences in many potentially important determinants of citations.

[Insert Table 1 about here]

First consider an author's experience. Rare cases of young geniuses making scientific breakthroughs (Simonton, 1988) notwithstanding, the number of citations an article receives generally increases with author experience (Jones, 2009). Experienced researchers have more know-how in doing research and have overcome competition to gain research support and publish their work, which gives them a potential set of readers who may cite their new work. The table shows that women scientists are less experienced than men in terms of years of work and number of publications.[15] Among *last* authors, women average 13.1 years of experience and 37.37 publications compared to men, who have 17.5 years of experience and 69.84 publications. Among *first* authors, women average 6.3 years of experience and 11.68 publications, compared to men, who have 8.9 years of experience and 24.44 publications. In ensuing regressions, we include an author's *experience,* measured in years, and *cumulative publications* since the author's first publication in the matched PubMed and MAG database.[16]

---

[14] Journal fixed effects account for the possibility that some journals may be more influential among authors and citing papers of a specific gender. This could be due to either the journal's focus or editorial board practices, though there is evidence that editors and referees are gender-neutral among leading journals in economics (Card et al., 2020).

[15] This is in part because women are still less likely to move up the academic career ladder (Huang et al., 2020).

[16] A left-censoring issue could arise when we track authors' entry year (first year of publication) and publication experience (cumulated publications and citations), since the first year of our matched data between MAG and

The past citations an author receives is another indicator of their potential to be cited in a new article. Male *last* authors average nearly twice as many cumulative citations by the publication year of the cited paper (2,898 citations) than female *last* authors (1,452 citations). Male *first* authors accumulate 815 citations, on average, compared to 333 citations for female *first* authors by the cited paper's publication year. Our regression model includes *cumulative citations* to control for authors' track records of impact.

Articles with more authors tend to receive more citations (Jones, 2009; Wuchty et al., 2007), in part because each author brings additional knowledge to the paper and in part because each author widens the network of potential citers. Articles with women *last* author have on average 0.33 fewer authors than articles with men *last* author, while articles with women *first* authors have 0.09 more authors than articles with men *first* authors. Our regression analyses include the *number of authors* in the article.

Longer articles may contain more knowledge than shorter ones, which should increase the number of forward citations. Measuring the length of an article by its *number of pages*, our data show that female *last* author and *first* author papers both have more authors per paper than male *last* author and *first* author papers. Though small, both gender differences are significant at the one percent level.

Longer reference lists may indicate more thorough research that could also produce more citations. Articles with women *last* authors make 0.31 fewer references on average than articles with men *last* authors, while articles with women *first* authors make 1.33 more references than men *first* authors. In our analysis, the length of the reference list is measured by the *number of references* the article makes.

Given the selectivity that high-impact journals have in publishing papers, the impact factor of the journal in which a paper appears is a likely indicator of the article's quality and thus of future citations. It also is a likely potential attractor of readers who might boost citations even of mediocre articles. In either case, impact factors are positively correlated with citations. Our data show that male-led articles average higher in journal impact factor than female-led articles, with an average journal impact factor among *last* authors of 3.49 for men compared to 3.07 for woman and among *first* authors of 3.47 for men compared to 3.23 for women.[17] We include *journal impact factor* in our regression.

PUBMED is 1975. About 5.7% of last authors and 1.9% of first authors in our baseline sample of cited articles published between 2002-2017 have entry year in or prior to 1975. For this group of individuals, we indeed do not know their exact year of entry and also miss some of their early-career publications. But due to its small representation in our sample, it is unlikely to significantly drive our results. Moreover, our cohort analyses revealed a consistent pattern when we exclude this left-censored cohort.

[17] Note that although we control for journal fixed effects in our specification, journal quality and "attractiveness" to forward citations could change over time and the longitudinal variations in journal impact factor help account for these changes. The *journal impact factor* is not readily available from PubMed nor MAG. We obtain this variable from the Reliance on Science dataset (Marx and Fuegi, 2020), which calculates the journal impact factor for year t as the number of times articles from years t-1 and t-2 are cited by other articles during year t, divided by the number of articles published during years t-1 and t-2.

Finally, women and men also tend to work in different fields and on different topics within a field, which may also impact citations. For instance, 10.1% of *last* authors in "Optics" are female, compared to 66.7% of *last* authors in "Nursing." We use MAG's hierarchical topic modelling to separate the gender homophily of citations within and across fields (Shen et al., 2018). The hierarchy distinguishes 288 fields of study, with an average of 8,447 articles per field in our sample. MAG's topic modelling algorithm also assigns multiple scientific concepts at varying degrees of refinement to each article and a probability score associated with each concept. We use the scientific concept with the highest probability score as the one that best describes an article. In this way, our sample is tagged to 59,411 distinct scientific concepts, averaging 41 articles per scientific concept. We include fixed effects for *fields* and *concepts* in our regression analyses.

4. Results

4.1 Forward Citations for Female-led and Male-led Articles

Table 2 gives summary statistics for the dependent variables on which we focus: overall citations, self-citations, citations from other male-led articles, and citations from other female-led articles. Consistent with earlier works (Caplar et al., 2017; Dworkin et al., 2020; Hutson, 2002; Long, 1992; Maliniak et al., 2013), our data show that women-led articles receive fewer forward citations than men-led articles both when we compare by *last* author gender (a gender differential of 1.93) and by *first* author gender (a gender differential of 1.06). Excluding self-citations, which advantage men because of their greater tendency to cite themselves and their greater number of articles from which to self-cite (Azoulay and Lynn, 2020; King et al., 2017; Mishra et al., 2018), the gender citation gap remains significant, and narrows to 1.57 fewer citations for women when compared by *last* author gender and 0.85 fewer citations for women when compared by *first* author gender.

If articles were assessed similarly by men and women, they should be recognized analogously by potential citing articles irrespective of gender. Following this reasoning, if men-led articles are of higher quality than women-led articles and thus deserve more citations, they should receive more citations from subsequent articles irrespective of whether these subsequent articles are led by men or women. However, as shown in Figures 2A and 2B, the data tell a different story, one in which forward citations exhibit gender homophily. We divide forward citations (*excluding* self-citations from now onward) into those from men- and women-led articles. Compared to women-led articles, men-led articles receive 1.02 (comparison by *first* author gender) to 1.59 *more* citations (comparison by *last* author gender) from subsequent men-led articles, but 0.20 (comparison by *last* author) to 0.30 *fewer* citations (comparison by *first* author gender) from subsequent women-led articles.

[Insert Table 2 and Figure 2 about here]

Finally, since the majority of citing articles are men-led articles, the number of citations from men-led articles to both genders is about 1.5 to 3 times that of the number of citations from women-led articles to both genders. Articles receive 6.03 citations on average from articles with male *last* authors and 2.12 citations from articles with female *last* authors, and receive 4.89 citations on average from articles with male *first* authors and 3.12 from article with female *first* authors. The advantage that women-led articles have in receiving citations from women-led articles is surpassed by their disadvantage in receiving citations from men-led articles, leading to a female disadvantage in total citations.

4.2 Effect of Observable Characteristics on Gender Homophily in Citations

To what extent, if at all, might the observed gender citation homophily in the previous section result from differences in the characteristics of female-led and male-led articles or the characteristics of authors? Table 3 answers this question in our data based on estimating regression equation (1). The controls for journal impact factor, author experience, author cumulated publications and citations, author affiliation, author team size, article length, and number of backward references, as well as year fixed effects, journal fixed effects, and field fixed effects, substantially reduce the estimated magnitude of gender citation homophily but still leave a significant relation where men-led articles receive a citation boost from men-led articles and women-led articles receive a citation boost from women-led articles.

Columns 4-6 in Table 3A and 3B show that, compared to men-led articles, women-led articles receive 0.13 (gendered by *last* author) to 0.15 (gendered by *first* author) more citations from women-led articles, and receive 0.42 (gendered by *last* author) to 0.33 (gendered by *first* author) fewer citations from men-led articles. Women-led articles' advantage in receiving forward citations from other women-led articles is unlikely to offset their disadvantage in receiving citations from other men-led articles, as women-led articles receive between 0.28 (gendered by *last* author) to 0.19 fewer citations (gendered by *first* author) than men-led articles. This accounts for 20.4% to 25.8% of the raw citation gap.[18]

[Insert Table 3 about here]

Even within the same field, men and women could work on different scientific concepts that would lead to women and men receiving citations from the other gender less frequently due to their lack of concept alignment. To account for the possibility of this sorting by gender, we add concept-level fixed effects in columns 7-9 of Table 3A and 3B. The concept definition is highly refined, with 59,411 distinct scientific concepts covering the 2.4 million articles in our analysis. The addition of this huge number of fixed effects invariably reduces the estimated effect of gender homophily on forward citations, but the sign and statistical significance remain, with an estimated disadvantage for women-led articles of 0.26

[18] This ratio is calculated as women's residual disadvantage in overall citations with observables and field fixed effects accounted for divided by the raw gender citation gap. It is 0.284/1.391=20.4% when gendered by last author and 0.187/0.726=25.8% when gendered by first author.

citations (gendered by *last* author) and 0.17 citations (gendered by *first* author). However, our findings at the field level are likely more meaningful for researchers' careers because when men and women are evaluated for funding or promotion purposes, they are usually compared to peers in a broader discipline than within a narrow group of researchers working on the same scientific concept.

4.3 Research Areas and Cohort Differences in Gender Homophily in Forward Citations

We now assess the extent to which gender citation homophily is found across fields and cohorts with varying gender compositions. In most of the 288 fields tagged by MAG with varying shares of women's representation, women-led articles receive a significantly higher percentage of citations from other women-led articles (and thus a lower percentage of citations from other men-led articles) compared to men-led articles. In Table 4, we report the top ten major fields for men and women[19] in life science and show gender differences in the composition of their forward citations. For instance, in the field of "Optics," men-led articles receive 13% of their forward citations from female-led articles, while women-led articles receive 15.9% of the forward citations from female-led articles, a 2.9% gap significant at the one percent level. In the field of "Nursing," 52.4% of the forward citations received by men-led articles are from female-led articles, while 66.1% of the forward citations received by women-led articles are from female-led articles, a 13.8% gap significant at the one percent level.

[Insert Table 4 about here]

To see whether and how gender homophily changes by gender ratio in research areas and the resulting effect on the gender gap with all other author and article characteristics accounted for, we classify each MAG scientific concept using the share of women's representation into categories of ten percentage points and run the same analyses as above for each subsample (Figure 3A and 3B, also see Appendix Table A2). We find that no matter the gender ratio, gender homophily persists. Compared to men-led articles, women-led articles receive significantly more citations from subsequent women-led articles (red line) and fewer citations from subsequent men-led articles (blue line). However, the gender gap in forward citations (grey bar) diminishes with higher shares of female-led authors in a scientific area – to the point where when the share of women authors is above par, the gender gap disappears, or it even becomes positive for women, though still insignificant.

[Insert Figure 3 about here]

This pattern implies that in research areas where women are underrepresented, women-led articles' advantage in receiving forward citations from women-led articles is unlikely to offset their disadvantage in receiving forward citations from men-led articles, thus resulting in a more sizable overall

[19] We rank research fields with at least 15,000 journal articles by the share of women last authors. Among these fields, we list the ten fields with the least share of women in Table 4A and the ten fields with the highest share of women in Table 4B.

gender citation gap. Such article-level citation gaps could accumulate over one's career and reinforce women's minority status. Meanwhile, in gender-balanced research areas, even when we observe no significant gender gap in citations, gender homophily in forward citation persists. Thus, when the ratio of women in certain research areas has attained gender equality that correlates with decreases in the gender gap of forward citations, it may still conceal gender-homophilic behaviors of authors in making citations. This leads to distorted knowledge flows, which further entrenches the gendered choice of research areas and limits the possibilities of combinatorial innovation that builds upon diversified prior knowledge.

Figure 1A shows an upward trend in the female share of life scientists from 2000 to 2020 that altered the gender demography of scientific work. To what extent, if at all, has gender homophily in citations and its impact on the gender citation gap declined for newer generations of scientists? We examine this question by dividing our sample into cohorts, using an author's year of first publication to place them into a specific group, from the oldest cohort, whose members published their first paper prior to 1985, to the newest cohort, whose members published their first paper between 2015 and 2017. We then estimate the extent of gender homophily for each of the eight cohort groups and graph the estimated coefficients in Figure 3C and 3D (also see Appendix Table A3). Each red dot shows the impact of having a female *last* or *first* author of a specific cohort on receiving citations from subsequent women-led articles. Each blue dot is the impact of having a female *last* or *first* author of a specific cohort on receiving citations from subsequent men-led articles. Although gender homophily persists (as indicated by the pattern that red lines lie above blue lines), it steadily diminishes for more recent cohorts of scientists (as red lines and blue lines get closer to each other). The gender gap (grey bars) also decreases, since women's participation in life sciences has increased over time. This pattern is consistent with the earlier field heterogeneity results. With improved women's representation in more recent cohorts, articles led by women are subject to a smaller disadvantage in garnering forward citations from men. Most of this disadvantage is offset by their advantage in receiving forward citations from other women, resulting in a smaller gender gap in overall forward citations.

4.4 Robustness Checks

In previous analyses, we used the gender of authors at key positions—i.e., as *first* and *last* author—to classify the gender of cited and citing articles. In this subsection, we provide additional robustness tests by altering the way we assign gender to articles. In Table 5, we classify a cited article's gender using different gender composition measures of the author team while still classifying citing articles using *last* (Table 5A) and *first* authors' gender (Table 5B). The dependent variables in all odd columns are citations made by women-led articles, while all even columns are citations made by men-led articles. In columns 1 and 2, we regress on the *female last (first) author* and the *number of female authors in non-last (non-first) positions*. In columns 3 and 4, we regress on *female last author*, *female first author*, and the *number of female authors in middle-author positions*. In columns 5 and 6, we regress on three indicator variables: (1) *minority female*, which equals 1 when female authors account for less than half of the authors in the team and zero otherwise; (2) *majority female*, which equals 1 when female authors account for half or more of the authors in the team and zero otherwise; and (3) *all female*, which equals 1 when all authors in the team are female and zero otherwise. In these two columns, articles with *all male* authors are the reference group; thus, its coefficient is omitted. Lastly, in columns 7 and 8, we simply regress on the *share of female*—i.e., the percentage of females in the author team. Our main finding of gender homophily in forward citations is robust to all these alternative specifications. The coefficients on all explanatory variables that proxy the share of female authors in an article are positive and significant on citations from women, as illustrated in odd model specifications. This shows that women tend to receive more citations from women. Conversely, the same coefficients of interest are all negative and significant on citations from men in even model specifications, which shows that men tend to receive more citations from men.

[Insert Table 5 about here]

We further decompose the forward citations received into the following: (1) the *number of citations from all male* articles, (2) the *number of citations from minority female* articles, (3) the *number of citations from majority female* articles, and (4) the *number of citations from all female* articles. In Table 6, we report regression results using these four variables as the dependent variable. Our main finding of gender homophily in forward citations is still robust, this time to various gender classifications for citers of an article. While controlling for the full set of observables and fixed effects, column 1 shows that compared to *all male* author teams, those in *minority female*, *majority female,* or *all female* teams receive significantly less citations from *all male* articles, thus showing again that men receive more citations from men. Conversely, columns 3 and 4 show that as compared to *all male* teams, those in *minority female*, *majority female,* or *all female* teams receive significantly more citations from *all female* articles or *majority female* articles, thus showing that women tend to receive more citations from women.

[Insert Table 6 about here]

We also experimented with alternative functional forms to the Table 3 regressions, repeating our main analyses using Poisson QML regression models. The results in Appendix Table A4 are consistent with our core findings.

In our control variables, aside from focusing on first and last author characteristics, the characteristics of middle authors, such as those affiliated with highly ranked institutions or with more research experience, can also affect the level of forward citations garnered by the articles. To address this concern, we construct eight new variables and add them as controls: (1) the *highest-ranked affiliation* among middle authors (ranked by affiliation-level number of publications in our sample) as fixed effect,

(2) the *share of middle authors from top 100 affiliations* (ranked by affiliation-level number of publications in our sample), (3) the *average years of experience* among middle authors, (4) the *average cumulated publications* among middle authors, (5) the *average cumulated citations* among middle authors, (6) the *longest years of experience* among middle authors, (7) the *highest cumulated publications* among middle authors, and (8) the *highest cumulated citations* among middle authors. In the analysis, we restrict our sample to articles with at least three authors so as to have middle authors. We report this set of robustness results in Appendix Table A5. More impactful middle authors with more prestigious affiliations indeed have positive effects on the citation counts, but our key variables of interest—i.e., the indicator for the gender of the first/last authors of the cited articles—remain robust to these added controls and fixed effects.

We repeat our analyses with an expanded sample of articles with non-U.S. first authors. We report these results in Appendix Table A6. As expected, in Eastern Asian countries where first name-based gender identification is very noisy, including South Korea, Japan, and China, we lack the statistical power to find a significant pattern of gender homophily (panel A). In the sample of articles with first authors from OECD countries minus the United States, South Korea, and Japan (panel B) and those with first authors from the "Rest of the world" (panel C), we find results consistent with our main findings (panel D) using our baseline sample of articles with U.S. first authors that we include for benchmarking.

Finally, we use two alternative sets of gender imputation criteria for first and last authors, one stricter and one laxer, and find robust and stable results throughout. The stricter criterion requires the frequency of a name in the Genderize.io database to be at least ten and the probability of the assigned gender to be at least 90%, which yields first and last author gender imputations of 69% for U.S.-based articles and 64% for non-U.S.-based articles. The laxer criterion does not impose any restriction on frequency or probability, and yields first and last author gender imputations of 94% for U.S.-based articles and 88% for non-U.S.-based articles. We report the robust results based on these different gender imputation criteria in Appendix Table A7.

In sum, our core findings are robust to different ways of assigning gender to articles, different samples, different characteristics of middle authors, and different econometric specifications.

5. Citation Homophily: Professional Connections, Discrimination Based on Gendered Names, or Specialized Interests?

Our findings above stipulate that gender homophily occurs in forward citations in life science, with adverse effects to women as a minority group in most fields. In this section, we examine three potential pathways that can contribute to this phenomenon: gendered professional connections, gender discrimination against unknown researchers based on gendered names, and gender-specialized research interests that are not captured in our measures of research areas.

5.1 Gender Homophily in Professional Connections

The first potential pathway to gender homophily in citations could be through professional connections that men and women maintain, such as who they collaborate, interact, and sit with at conferences (Chai and Freeman, 2019). These interactions influence how knowledge diffuses and even scientists' choice of research focus (Gazni and Thelwall, 2014; Wallace et al., 2012). Close professional acquaintances – both those linked informally and formal collaborators – know each other's work better and are therefore more likely sources of citations. Works of collaborators or friends may also receive one's citations as a favor even when only remotely related. Meanwhile, the composition of researchers' professional connections is also likely to be gender-homophilic. Building on the likelihood that people are more likely to attract citations from their professional connections and that those connections exhibit gender homophily, we posit that gender homophily in researchers' past professional connections can be a contributing factor to the gender homophily observed in forward citations in life sciences.

While it is difficult to measure both formal and informal professional connections in general, their gender composition is likely to be correlated with the gender composition of collaborators. We use direct collaborators as a proxy for professional connections by including researchers who have published a paper with one or more authors of a cited article's author team in the three years prior to the article's publication. For example, we identify collaborators for articles published in 2005 using publications from 2002-2004. Thus, in this analysis, we take articles published from 2005 (instead of 2002 in the main sample) to 2017 as our sample. Table 7 compares the numbers of male, female, and overall collaborators by the gender of the cited article's *last* author (in Table 7A) and *first* author (in Table 7B). Consistent with prior findings (Holman and Morandin, 2019; Lee et al., 2019), we observe that the collaborator composition also exhibits gender homophily: both genders have a higher share of same-gender collaborators than predicted by randomness. Women make up 35.0% of collaborators in articles of our sample, on average. However, the share of women collaborators among articles with a male *last* author is lower, at 32.4%, and that among articles with a female *last* author is higher, at 42.6% – a 10.3% difference significant at the one percent level. Similarly, the average share of female collaborators among articles with a male *first* author is 31.8%, while that among those with a female *first* author is 40.6% – an 8.8% difference significant at the one percent level.

[Insert Table 7 about here]

To show how collaborator composition affects forward citations, we first replicate the original regression specification in equation (1) with the new sample[20] and report these results in columns 1-3 of Table 8 to provide a benchmark for comparison. We then add the *number of male collaborators* and

[20] Since we used a three-year window prior to the publication year of the cited article to construct variables for collaborators, the sample used in Table 8 is smaller than the baseline sample.

number of female collaborators in equation (1) and report the regression results in column 4-6. Gender citation homophily also shows up in collaborators' gender composition. Specifically, in Table 8A, where we use *last* author gender to classify both the cited and citing articles, columns 4 and 5 show that an increase of 50 in the number of male collaborators decreases the number of citations from female-led articles by 0.05 but increases the number of citations from male-led articles by 0.60.[21] Meanwhile, an increase of 50 in the number of female collaborators increases the number of citations from female-led articles by 0.20 but decreases the number of citations from male-led articles by 0.65. Taken jointly, we observe in column 6 that increasing the *number of male collaborators* increases total citations, whereas increasing the *number of female collaborators* does the opposite.

Although these changes in effect sizes are small, it's important to note that these variables are noisy proxies for the size and composition of author teams' professional connections. After all, collaborators within a three-year window prior to the publication year of a cited article are a subset of the author team's collaborative connections, which is itself a subset of their professional connections. In Table 8B, we find consistent patterns when we use *first* author gender to classify both the cited and citing articles. Lastly, comparing columns 4 and 1, columns 5 and 2, and columns 6 and 3, the coefficient size of the variable *female last author* decreases when we include the *number of male collaborators* and *number of female collaborators* in regressions. The direction of these changes supports the notion that part of the observed gender citation homophily operates through gender homophily in professional connections.

[Insert Table 8 about here]

5.2 Discrimination Based on Gendered Names

Let's now assume a polar opposite situation to researchers citing articles written by authors they already know and consider one where researchers randomly come across relevant works written by authors they do not know. When deciding whether or not to cite these works, potential citers could place greater faith in articles written by someone of the same gender by inferring the author's gender using first names and base their referencing decision on that inference. We test for this "discrimination based on gendered name" hypothesis by taking advantage of variations in the inclusion of full first names in journal articles – for example, cases in which Jennifer Smith's name was given in full compared to cases where it was listed as J. Smith. The hypothesis implies that, among articles led by Jennifer Smith, those on which she is listed as J. Smith will receive more citations from male-led articles (and fewer citations from female-led articles) than those on which she is listed as Jennifer Smith.

To test whether this is the case, we estimate:

---

[21] The mean, mode, and standard deviation of the *number of male collaborators* are 81, 38 and 129 respectively. The mean, mode, and standard deviation of the *number of female collaborators* are 44, 20, and 76 respectively. The regression coefficients reported in Table 8 show changes in citations associated with an increase in the *number of male (or female) collaborators* by 1.

$$citations_{itjfs}^{G} = \beta_0 + \beta_1 OnlyInitials_{is} + \beta_2 female_s * OnlyInitials_{is} + \Gamma X_{i,j,t} + \lambda_j + \sigma_f + \mu_t + \phi_s + \epsilon_{itjfs} \quad (2)$$

using a subsample consisting of 797,382 articles written by *last* authors who sometimes reported full first names and sometimes only reported first initials, but whose gender we have imputed from our disambiguated author database. Starting 2002,[22] PubMed began recording authors' first initials or full first names exactly as they appear at the designated author position in the article, which makes it possible to observe a researcher's distinctive name-reporting practices across articles. The variable $OnlyInitials_{is}$ equals 1 when individual s does not report his or her full first name on article i, where the individual is listed as *last* author. We also control for a series of individual fixed effects (ϕ_s). Note that the gender indicator of individual s ($female_s$) is absorbed by the individual fixed effects and does not appear in the regression on its own.

We are interested in β_2, the coefficient of the interaction term between $OnlyInitial_{is}$ and the gender indicator $female_s$. For example, when the dependent variable is the number of forward citations from male-led articles, an insignificant β_2 indicates that, although women-led articles are at a disadvantage in receiving citations from male-led articles, hiding their gender by not reporting their full first name does not alter such a disadvantage. We also perform the same analysis with a subsample consisting of 381,957 articles written by *first* authors who occasionally do not report their full first names.

We report these regression results in Table 9. Specifically, for the same female author, the number of forward citations she receives from female-led articles does not significantly vary whether she reports her full name or not (columns 1 and 4). Similarly, for the same female author, the number of forward citations she receives from male-led articles also does not significantly vary whether she reports her full name or not (columns 2 and 5). These results suggest that whether or not she reports her first name (i.e., whether the gender could be guessed by strangers or not) does not change the citations to articles led by this author from both male- and female-led articles. These results help us rule out the "discrimination based on gendered name" hypothesis, as we do not observe gender citation homophily to be driven by discrimination on the basis of gender inferred from first names.

[Insert Table 9 about here]

5.3 Gender-Specialized Research Interests

The third possible pathway to gender citation homophily is gender-specialized research interests not captured in our measures of research areas. In our baseline results, we compared forward citations to men- and women-led articles within the same field and scientific concept. It is possible, nevertheless, that even among researchers who study the same scientific concept, men and women could still sort into

[22] Prior to 2002, PubMed only recorded authors' first initials.

different narrower topics, resulting in "topic homophily" in citations (citing articles that address the relevant topic) that appears as gender homophily (citing articles written by the same gender). The literature has indeed found that men and women sort into different research topics and that topics have different attractiveness to future citations. For example, Kozlowski et al. (2022, p. 6) find that "minoritized groups are overrepresented in lowly cited topics and underrepresented in highly cited topics; and that their work is less cited within and across topics, especially where they are underrepresented."

To further address gendered sorting across research topics, we leverage the PubMed Related Citations Algorithm (PMRA) (Lin and Wilbur, 2007). For every article available on PubMed, this algorithm identifies a ranked list of closest papers in idea space using topic-modelling techniques performed on Medical Subject Heading (MeSH) keywords.[23] We match each cited article in our sample with the article most similar to it according to the PMRA among those published in the same year and with a U.S.-based first author (Azoulay et al., 2015; Marx and Hsu, 2022).

We then compare citations within each article pair and report the results in Table 10.[24] In Table 10A, we include the results with both cited and citing articles gendered according to *last* author. In columns 1-3, we first replicate our original specification with the matched sample as benchmark for comparison. In columns 4-6, we add the matched article pair fixed effects, with each cited article and its match sharing a fixed effect. The coefficients on gender dummies thus reflect the within-pair difference. Using citations from female-led articles as the dependent variable, the coefficient on *female last author* in column 4 remains positive but is much smaller than that in column 1 and no longer statistically significant. This implies that matched articles with one being male-led and the other female-led are treated the same by female-led articles. Using citations from male-led articles, the coefficient on *female last author* in column 5 is still negative and statistically significant, though the effect size is approximately half of that in column 2. In Table 10B, we include results with both cited and citing articles gendered according to *first* author, and the patterns are similar.

[Insert Table 10 about here]

These results suggest that gendered sorting into narrow topics accounts for part of the gender citation homophily observed at the field or scientific concept level. It is worth noting, however, that scientists' choice of research topics could be a result of the knowledge diffusion process. Gender-biased knowledge diffusion and gender sorting could reinforce each other. Again, since scientists are usually

[23] MeSH keywords are independently indexed keywords performed by indexers of the United States National Library of Medicine at the National Institutes of Health.

[24] Note that only for a more limited set of articles (1,858,952 focal articles and 1,858,952 matched articles) we can find their related articles with US affiliated authors published in the same year. Moreover, the PMRA does not assure symmetric matching, in that for focal article A we may find article B to be its most similar article, but when we treat article B as the focal article and search for its similar articles, a third article C could be more similar to B than A is.

evaluated and compared to peers in a broader field rather than a refined topic in promotions and funding decisions, the gender citation homophily and the gender citation gap at the field level are more relevant to both scientists' careers and science policies.

6. Discussion & Conclusion

Taken together, our findings demonstrate that forward citations exhibit gender homophily broadly, irrespective of the share of women in the field. Given that the majority of life scientists are men and that men publish more than women (Ceci et al., 2014; Lerchenmueller and Sorenson, 2018; Xie and Shauman, 1998), gender homophily leaves women at a detriment in garnering forward citations. Citations are not the objective measure of the scientific merit of articles that some may view them to be. Other factors also influence citations in ways that systematically disadvantage women in hiring, promotion, and funding decisions that rely on publication and citation counts. Gender homophily in citation attributions, just like gender homophily in authorship attributions (Ross et al., 2022), harms women's career advancement in academic research when all else equal.

Fortunately for science and women scientists in particular, a trend toward increases in female representation in the life sciences has begun to alleviate the adverse effects of gender homophily. As shown in our heterogeneity analyses, more recent cohorts of scientists are less subject to gendered preferences in receiving citations (Figure 3C and 3D), and higher shares of female-led authors in a scientific field are associated with a smaller gender gap in forward citations. And, as noted, when the share of women authors is above par, the gender gap becomes positive, though insignificant for women (Figure 3A and 3B).

Still, gender homophily remains and distorts the flow of knowledge and credit for contributions to knowledge. To the extent that science advances by climbing "on the shoulders of giants" (Newton, 1676), our findings suggest that it has been too much on the shoulders of *similar* gendered giants, contributing to gendered choices in research focus (Kozlowski Diego et al., 2022), which more likely than not retard the advance of science, as gender-diverse teams tend to produce higher impact and more novel research (Yang et al., 2022).

Although we have made use of rich bibliometric data to document a robust pattern of gender homophily in citations, our paper is limited in pinpointing the ultimate mechanism driving such patterns, especially given the complex relationship between the gendered choice of research topics and gender homophily in professional networks. To the fronts of both advancing this literature and designing policy interventions, experimenting with increasing gender-neutral interactions at conferences to widen networks would appear to be a good starting point. Organizers could design seating plans to better mix people and reduce gender or re-occurring groupings. Randomly assigned seat numbers would place more women and men next to each other unobtrusively at dinners and discussions. Moreover, the effectiveness of interventions that seek to widen gender-based networks could be tested at low costs. Prior literature using quasi-natural experiments (Chai and Freeman, 2019) and randomized controlled experiments (Boudreau et al., 2017) has shown that temporary collocation increases collaboration. Conference organizers and researchers could design experiments that fit with the group to find out what works. Follow-up interviews or surveys could get scientists' reactions and ideas of what might work better, and follow-up bibliometric analysis could test if an initiative leads to a more gender-inclusive citing practice, more gender-balanced co-authorships, and a more gender-integrated choice of research interest.

Many scientific societies have developed women-specific associations or women-targeted committees to promote women-to-women networking and to encourage senior women to mentor junior women and improve their career prospects, such as the Women Chemists Committee of the American Chemists Society (ACSWCC) and the Committee on the Status of Women in Physics of the American Physics Society (CSWP). However, more needs to be done to create a more gender-integrated scientific community. It will be helpful to raise awareness on the existence and consequences of gender homophily among scientists and policymakers. People's default networking behavior could make them more likely to miss good ideas for their own work from those outside their gender-biased network, while inadvertently opening the door for the citation homophily that harms women's careers. Aware of such unconscious biases, individual researchers could step outside their network to be more gender-inclusive in professional communications as well as when searching for literature and ideas. Policymakers and administrators could also factor in gender citation homophily when evaluating researchers and when enacting diversity, equity, and inclusion policies (Graddy-Reed and Lanahan, 2023). Joint efforts from all stakeholders are needed to move toward a more gender-integrated scientific community that would create both fairer opportunities for women and better science by enhancing diversity in how knowledge diffuses and is subsequently recombined (Fleming, 2001; Weitzman, 1998).

References

Azoulay, P., Furman, J.L., Krieger, Murray, F., 2015. Retractions. The Review of Economics and Statistics 97, 1118–1136. https://doi.org/10.1162/REST_a_00469

Azoulay, P., Lynn, F.B., 2020. Self-citation, cumulative advantage, and gender inequality in science. Sociological science 7, 152–186.

Baerlocher, M.O., Newton, M., Gautam, T., Tomlinson, G., Detsky, A.S., 2007. The Meaning of Author Order in Medical Research. Journal of Investigative Medicine 55, 174–180. https://doi.org/10.2310/6650.2007.06044

Beaudry, C., Larivière, V., 2016. Which gender gap? Factors affecting researchers' scientific impact in science and medicine. Research Policy 45, 1790–1817. https://doi.org/10.1016/j.respol.2016.05.009

Bhattacharya, J., Packalen, M., 2020. Stagnation and scientific incentives. National Bureau of Economic Research.

Bikard, M., Fernandez-Mateo, I., 2022. Standing on the Shoulders of (Male) Giants: Gender Inequality and the Technological Impact of Scientific Ideas. Available at SSRN 4059813.

Bol, T., de Vaan, M., van de Rijt, A., 2022. Gender-equal funding rates conceal unequal evaluations. Research Policy 51, 104399. https://doi.org/10.1016/j.respol.2021.104399

Boudreau, K., Ganguli, I., Gaule, P., Guinan, E., Lakhani, K., 2017. A field experiment on search costs and the formation of scientific collaborations. Review of Economics & Statistics.

Caplar, N., Tacchella, S., Birrer, S., 2017. Quantitative evaluation of gender bias in astronomical publications from citation counts. Nature Astronomy 1, 0141. https://doi.org/10.1038/s41550-017-0141

Card, D., DellaVigna, S., Funk, P., Iriberri, N., 2020. Are Referees and Editors in Economics Gender Neutral?*. The Quarterly Journal of Economics 135, 269–327. https://doi.org/10.1093/qje/qjz035

Ceci, S.J., Ginther, D.K., Kahn, S., Williams, W.M., 2014. Women in Academic Science: A Changing Landscape. Psychol Sci Public Interest 15, 75–141. https://doi.org/10.1177/1529100614541236

Ceci, S.J., Williams, W.M., 2011. Understanding current causes of women's underrepresentation in science. Proc Natl Acad Sci USA 108, 3157. https://doi.org/10.1073/pnas.1014871108

Chai, S., D'Amour, A., Fleming, L., 2020. Explaining and predicting the impact of authors within a community: an assessment of the bibliometric literature and application of machine learning. Industrial and Corporate Change 29, 61–80. https://doi.org/10.1093/icc/dtz042

Chai, S., Freeman, R.B., 2019. Temporary colocation and collaborative discovery: Who confers at conferences. Strategic Management Journal 40, 2138–2164.

Chatterjee, P., Werner, R.M., 2021. Gender Disparity in Citations in High-Impact Journal Articles. JAMA network open 4, e2114509. https://doi.org/10.1001/jamanetworkopen.2021.14509

Cole, J.R., Zuckerman, H., 1984. The Productivity Puzzle: Persistence and Change in Patterns of Publication Among Men and Women Scientists In: Steinkamp, MW, Maehr, M.(Eds.): Advances in Motivation and Achievement.

Dion, M.L., Sumner, J.L., Mitchell, S.M., 2018. Gendered Citation Patterns across Political Science and Social Science Methodology Fields. Political Analysis 26, 312–327.

Dworkin, J.D., Linn, K.A., Teich, E.G., Zurn, P., Shinohara, R.T., Bassett, D.S., 2020. The extent and drivers of gender imbalance in neuroscience reference lists. Nature Neuroscience 23, 918–926. https://doi.org/10.1038/s41593-020-0658-y

Ellison, G., 2013. How does the market use citation data? The Hirsch index in economics. American Economic Journal: Applied Economics 5, 63–90.

Ertug, G., Gargiulo, M., Galunic, C., Zou, T., 2018. Homophily and Individual Performance. Organization Science 29, 912–930. https://doi.org/10.1287/orsc.2018.1208

Fleming, L., 2001. Recombinant uncertainty in technological search. Management Science 47, 117–132.

Freeman, R.B., Huang, W., 2015. Collaborating with people like me: Ethnic coauthorship within the United States. Journal of Labor Economics 33, S289–S318.

Gazni, A., Thelwall, M., 2014. The long-term influence of collaboration on citation patterns. Research Evaluation 23, 261–271. https://doi.org/10.1093/reseval/rvu014

Ghiasi, G., Mongeon, P., Sugimoto, C., Larivière, V., 2018. Gender homophily in citations, in: STI 2018 Conference Proceedings. Centre for Science and Technology Studies (CWTS), pp. 1519–1525.

Gibson, C.B., Zellmer-Bruhn, M.E., 2001. Metaphors and Meaning: An Intercultural Analysis of the Concept of Teamwork. Administrative Science Quarterly 46, 274–303. https://doi.org/10.2307/2667088

Graddy-Reed, A., Lanahan, L., 2023. Prioritizing diversity? The allocation of US federal R&D funding. Science and Public Policy 50, 104–119. https://doi.org/10.1093/scipol/scac052

Graddy-Reed, A., Lanahan, L., Eyer, J., 2019. Gender discrepancies in publication productivity of high-performing life science graduate students. Research Policy 48, 103838. https://doi.org/10.1016/j.respol.2019.103838

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proceedings of the National academy of Sciences 102, 16569–16572.

Hofstra, B., Kulkarni, V.V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., McFarland, D.A., 2020. The Diversity–Innovation Paradox in Science. Proceedings of the National Academy of Sciences 117, 9284–9291. https://doi.org/10.1073/pnas.1915378117

Holman, L., Morandin, C., 2019. Researchers collaborate with same-gendered colleagues more often than expected across the life sciences. PLoS One 14, e0216128–e0216128. https://doi.org/10.1371/journal.pone.0216128

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

Hutson, S.R., 2002. Gendered Citation Practices in American Antiquity and Other Archaeology Journals. American Antiquity 67, 331–342. https://doi.org/10.2307/2694570

Isaacson, W., 2021. The code breaker : Jennifer Doudna, gene editing, and the future of the human race, First Simon&Schuster hardcover edition. ed. Simon & Schuster, New York.

Jones, B.F., 2009. The burden of knowledge and the "death of the Renaissance Man": Is innovation hetting harder? Review of Economic Studies 76, 283–317.

Joshi, A., Labianca, G., Caligiuri, P.M., 2002. Getting along long distance: understanding conflict in a multinational team through network analysis. Journal of World Business 37, 277–284. https://doi.org/10.1016/S1090-9516(02)00094-9

King, M.M., Bergstrom, C.T., Correll, S.J., Jacquet, J., West, J.D., 2017. Men Set Their Own Cites High: Gender and Self-citation across Fields and over Time. Socius 3, 2378023117738903. https://doi.org/10.1177/2378023117738903

Klug, A., 1968. Rosalind Franklin and the Discovery of the Structure of DNA. Nature 219, 808–810. https://doi.org/10.1038/219808a0

Koning, R., Samila, S., Ferguson, J.-P., 2021. Who do we invent for? Patents by women focus more on women's health, but few women get to invent. Science 372, 1345–1348. https://doi.org/10.1126/science.aba6990

Kozlowski, D., Larivière, V., Sugimoto, C.R., Monroe-White, T., 2022. Intersectional inequalities in science. Proceedings of the National Academy of Sciences 119, e2113067119. https://doi.org/10.1073/pnas.2113067119

Kozlowski Diego, Larivière Vincent, Sugimoto Cassidy R., Monroe-White Thema, 2022. Intersectional inequalities in science. Proceedings of the National Academy of Sciences 119, e2113067119. https://doi.org/10.1073/pnas.2113067119

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

Lazarsfeld, P.F., Merton, R.K., 1954. Friendship as a social process: A substantive and methodological analysis. Freedom and control in modern society 18, 18–66.

Lee, E., Karimi, F., Wagner, C., Jo, H.-H., Strohmaier, M., Galesic, M., 2019. Homophily and minority-group size explain perception biases in social networks. Nature Human Behaviour 3, 1078–1087. https://doi.org/10.1038/s41562-019-0677-4

Lerchenmueller, M.J., Sorenson, O., 2018. The gender gap in early career transitions in the life sciences. Research Policy 47, 1007–1017. https://doi.org/10.1016/j.respol.2018.02.009

Levin, S.G., Stephan, P.E., 1998. Gender Differences in the Rewards to Publishing in Academe: Science in the 1970s. Sex Roles 38, 1049–1064.

Lin, J., Wilbur, W.J., 2007. PubMed related articles: a probabilistic topic-based model for content similarity. BMC Bioinformatics 8, 423. https://doi.org/10.1186/1471-2105-8-423

Long, J.S., 1992. Measures of Sex Differences in Scientific Productivity*. Social Forces 71, 159–178. https://doi.org/10.1093/sf/71.1.159

Maliniak, D., Powers, R., Walter, B.F., 2013. The gender citation gap in international relations. International Organization 67, 889–922.

Marx, M., Fuegi, A., 2020. Reliance on science: Worldwide front-page patent citations to scientific articles. Strategic Management Journal 41, 1572–1594. https://doi.org/10.1002/smj.3145

Marx, M., Hsu, D.H., 2022. Revisiting the Entrepreneurial Commercialization of Academic Science: Evidence from "Twin" Discoveries. Management Science 68, 1330–1352. https://doi.org/10.1287/mnsc.2021.3966

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. Annual review of sociology 27, 415–444.
Mishra, S., Fegley, B.D., Diesner, J., Torvik, V.I., 2018. Self-citation is the hallmark of productive authors, of any gender. PLOS ONE 13, e0195773. https://doi.org/10.1371/journal.pone.0195773
Montpetit, É., Blais, A., Foucault, M., 2008. What Does it Take for a Canadian Political Scientist to be Cited?*. Social Science Quarterly 89, 802–816. https://doi.org/10.1111/j.1540-6237.2008.00561.x
Newton, I., 1676. Letter to Robert Hooke.
Reagans, R., 2011. Close Encounters: Analyzing How Social Similarity and Propinquity Contribute to Strong Network Connections. Organization Science 22, 835–849. https://doi.org/10.1287/orsc.1100.0587
Ross, M.B., Glennon, B.M., Murciano-Goroff, R., Berkes, E.G., Weinberg, B.A., Lane, J.I., 2022. Women are Credited Less in Science than are Men. Nature. https://doi.org/10.1038/s41586-022-04966-w
Santamaría, L., Mihaljević, H., 2018. Comparison and benchmark of name-to-gender inference services. PeerJ Computer Science 4, e156.
Shen, Z., Ma, H., Wang, K., 2018. A web-scale system for scientific knowledge exploration. arXiv preprint arXiv:1805.12216.
Simonton, D.K., 1988. Age and outstanding achievement: What do we know after a century of research? Psychological Bulletin 104, 251.
Teich, E.G., Kim, J.Z., Lynn, C.W., Simon, S.C., Klishin, A.A., Szymula, K.P., Srivastava, P., Bassett, L.C., Zurn, P., Dworkin, J.D., Bassett, D.S., 2022. Citation inequity and gendered citation practices in contemporary physics. Nature Physics 18, 1161–1170. https://doi.org/10.1038/s41567-022-01770-1
Van Raan, A.F., 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. scientometrics 67, 491–502.
Vásárhelyi, O., Zakhlebin, I., Milojević, S., Horvát, E.-Á., 2021. Gender inequities in the online dissemination of scholars' work. Proceedings of the National Academy of Sciences 118, e2102945118. https://doi.org/10.1073/pnas.2102945118
Wallace, M.L., Larivière, V., Gingras, Y., 2012. A Small World of Citations? The Influence of Collaboration Networks on Citation Practices. PLOS ONE 7, e33339. https://doi.org/10.1371/journal.pone.0033339
Wang, K., Shen, Z., Huang, C., Wu, C.-H., Eide, D., Dong, Y., Qian, J., Kanakia, A., Chen, A., Rogahn, R., 2019. A review of microsoft academic services for science of science studies. Frontiers in Big Data 2, 45.
Wang, X., Dworkin, J.D., Zhou, D., Stiso, J., Falk, E.B., Bassett, D.S., Zurn, P., Lydon-Staley, D.M., 2021. Gendered citation practices in the field of communication. Annals of the International Communication Association 45, 134–153. https://doi.org/10.1080/23808985.2021.1960180
Weitzman, M.L., 1998. Recombinant Growth. The Quarterly Journal of Economics 113, 331–360.
Wuchty, S., Jones, B.F., Uzzi, B., 2007. The Increasing Dominance of Teams in Production of Knowledge. Science 316, 1036–1039.
Xie, Y., Shauman, K.A., 2003. Women in science : career processes and outcomes. Harvard University Press, Cambridge, Mass.
Xie, Y., Shauman, K.A., 1998. Sex Differences in Research Productivity: New Evidence about an Old Puzzle. American Sociological Review 63, 847–870. https://doi.org/10.2307/2657505
Yang, Y., Tian, T.Y., Woodruff, T.K., Jones, B.F., Uzzi, B., 2022. Gender-diverse teams produce more novel and higher-impact scientific ideas. Proceedings of the National Academy of Sciences 119, e2200841119. https://doi.org/10.1073/pnas.2200841119

A. Share of women among life scientists **B. Gender gap in forward citations**

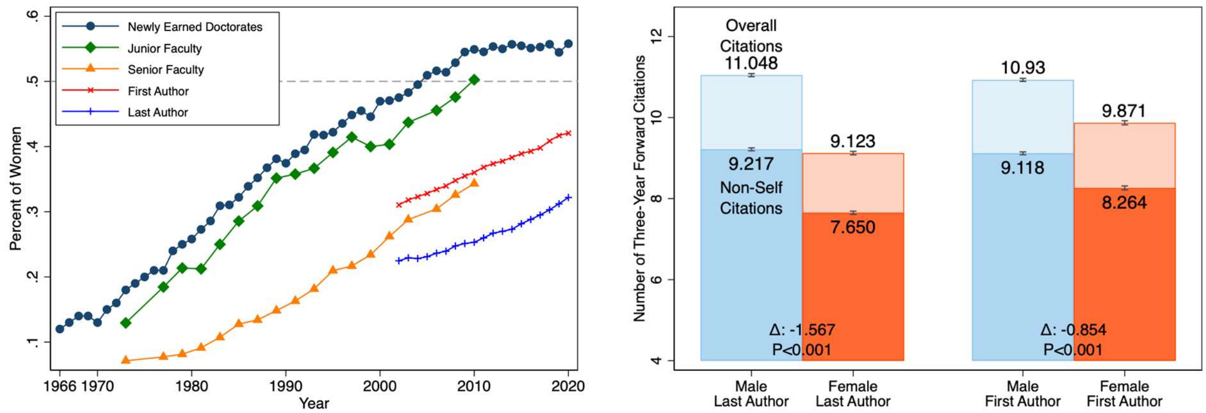

Figure 1: Women's representation in life sciences and the gender citation gap

Note: **(A)** Female share among newly earned doctorates, junior faculty, and senior faculty in the field of biological, agricultural, environmental, and health sciences; the share of female first author and the share of female last author among U.S.-based PubMed articles. The share of women among earned doctorates in life sciences by degree year (1966-2020) comes from the SED.[25] The share of women among doctorate holders employed in academia as full-time junior faculty and those employed in academia as full-time senior faculty during 1973-2010 are from NSF Science and Engineering (NSF S&E) Indicators 2014, Table 5-15.[26] The last year for which this tabulation is available is 2014. Later versions of NSF S&E indicators no longer provide this tabulation. According to the tabulation notes, academic employment is limited to U.S. doctorate holders employed at two- or four-year colleges or universities. Full-time senior faculty includes full professors and associate professors. Full-time junior faculty includes assistant professors and instructors from 1973 to 1995; from 1997 to 2010, full-time junior faculty includes assistant professors. The share of articles with female first authors and those with female last authors among U.S.-based PubMed articles are calculated from the same sample construction for all analyses, as elaborated in Section 3.
(B) Forward citations received by cited article gender. Cited article gender is classified by *last* author for the left panel and *first* author for the right panel. Wherever visible, error bars denote 95% confidence intervals. Self-citations are measured as forward citations made by any author in the author team of the cited article. Δ indicates the gender gap (female-male) in citations, excluding self-citations.

[25] Available here: https://ncsesdata.nsf.gov/home.
[26] Available here: http://www.nsf.gov/statistics/seind14/index.cfm/appendix/tables.htm#c3.

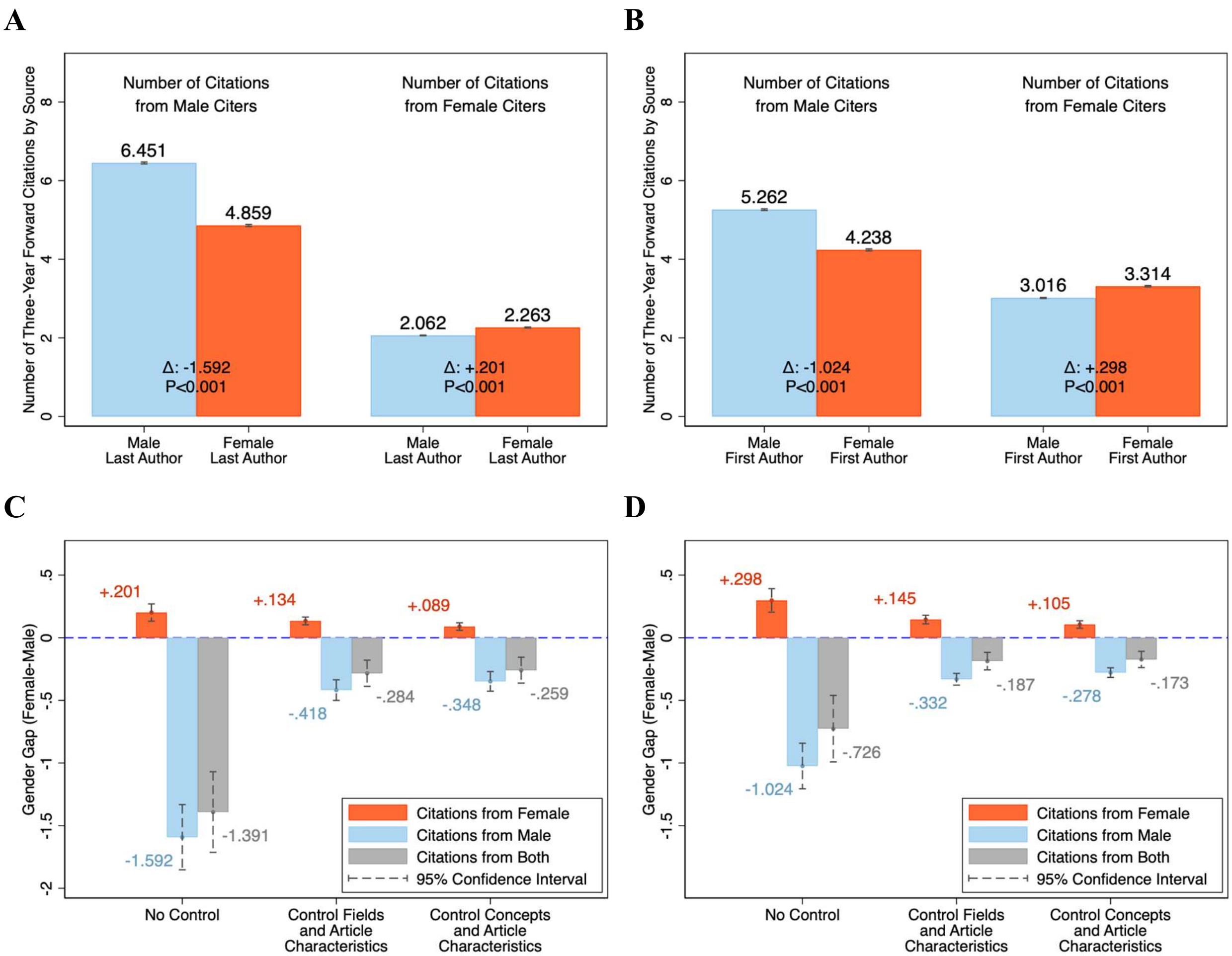

Figure 2: Gender homophily in forward citations

Notes: **(A) & (B)** Forward citations (from here onward, all forward citations refer to citations excluding self-citations) received by *female* and *male*-led cited article from *female* and *male*-led citing articles; citing and cited article gender are classified by *last* author (A) and *first* author (B).

(C) & (D) Coefficients for forward citations received from *male*- and *female*-led articles regressed on the dummy indicating that the cited article is written by *female*; cited and citing article gender are classified by *last* author (C) and *first* author (D). Women's disadvantage in garnering citations (the grey bar) is equivalent to women's disadvantage in citations from male citers (the blue bar) subtracted by women's advantage in citations from female citers (the orange bar).

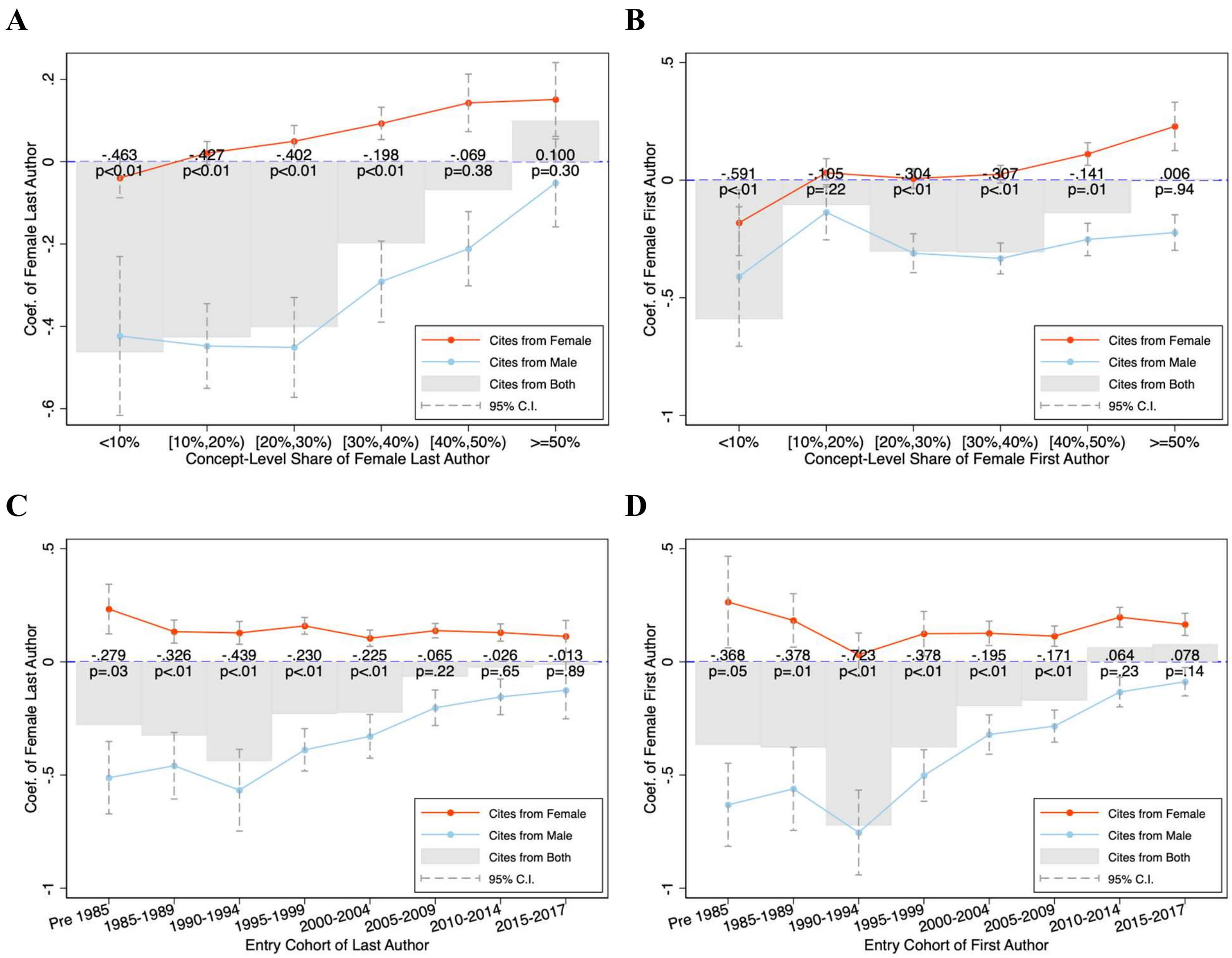

Figure 3: Heterogeneity in gender homophily and gender gap in forward citations
Notes: In all four graphs, the grey bars represent the gender gap in overall citations from female- and male-led articles for each split sample, and its magnitude is specified by the number above the zero-line, while the p-value is underneath.
(A) & (B) Split sample analysis by share of female authors in scientific concept. Each point indicates a subsample coefficient from regressing forward citations received from *male-* or *female-led citing articles* on the dummy indicating that the cited article is written by *female* and a complete set of controls. Cited article and citing article gender are classified by *last* author (A) and *first* author (B).
(C) & (D) Split sample analysis by authors' entry cohorts as defined by authors' first year of publication. Each point indicates a subsample coefficient from regressing forward citations received from *male* or *female-led citing articles* on the dummy indicating that the cited article is written by *female* and a complete set of controls. Cited article and citing article gender are classified by *last* author (C) and *first* author (D).

Table 1. Gender differences in article and author characteristics

(A)	*Last Author Gender*		Gap	
	Male	Female	Female-Male	
Number of Articles	1,795,497	637,309		
	73.80%	26.20%		
Article Characteristics				
Number of References	30.209	29.903	-0.305	***
	(31.820)	(30.928)		
Number of Pages	8.121	8.549	0.428	***
	(5.311)	(5.510)		
Journal Impact Factor	3.494	3.068	-0.426	***
	(3.190)	(2.866)		
Author Characteristics				
Author Team Size	4.629	4.303	-0.327	***
	(2.961)	(2.820)		
Last Author Experience	17.477	13.066	-4.411	***
	(10.658)	(10.346)		
Last Author Cumulated Publications	69.841	37.370	-32.472	***
	(94.007)	(55.435)		
Last Author Cumulated Citations	2898.031	1452.421	1445.610	***
	(6502.366)	(3571.737)		
Number of Female at Non-Last Position	1.124	1.443	0.319	***
	(1.387)	(1.573)		
Share with Female First Author+	33.0%	49.5%	16.5%	***
First Author Experience	6.969	7.687	0.718	***
	(8.232)	(8.881)		
First Author Cumulated Publications	16.480	17.509	1.029	***
	(41.694)	(42.882)		
First Author Cumulated Citations	491.957	529.857	-37.900	***
	(2251.695)	(2375.033)		

(B)	First Author Gender		Gap	
	Male	Female	Female-Male	
Number of Articles	1,544,727	888,079		
	63.50%	36.50%		
Article Characteristics				
Number of References	29.644	30.971	1.327	***
	(32.119)	(30.628)		
Number of Pages	8.074	8.510	0.436	***
	(5.424)	(5.256)		
Journal Impact Factor	3.470	3.231	-0.239	***
	(3.219)	(2.915)		
Author Characteristics				
Author Team Size	4.510	4.603	0.092	***
	(2.955)	(2.881)		
First Author Experience	8.862	6.283	-2.579	***
	(9.695)	(7.631)		
First Author Cumulated Publications	24.438	11.682	-12.756	***
	(56.201)	(27.349)		
First Author Cumulated Citations	814.944	333.277	481.667	***
	(3312.8)	(1458.685)		
Number of Female at Non-First Position	0.944	1.385	0.442	***
	(1.288)	(1.542)		
Share with Female Last Author+	20.7%	34.2%	13.5%	***
Last Author Experience	16.401	16.735	0.334	***
	(10.614)	(10.653)		
Last Author Cumulated Publications	63.621	62.222	-1.399	***
	(88.429)	(85.661)		
Last Author Cumulated Citations	2603.553	2591.878	11.675	0.163
	(6015.37)	(5889.603)		

Note: *** $p < 0.01$ from two-sided t-test. We classify cited articles by *last* author for panel (A) and by *first* author gender for panel (B). + For articles with at least two authors.

Table 2. Gender differences in forward citations by source of citation

(A)	Last Author		
	Male	Female	Female-Male
Number of Articles	1,795,497	637,309	
	73.8%	26.2%	
Number of Forward Citations, Three-Year Window:			
All	11.048	9.123	-1.925 ***
	(25.796)	(17.552)	
Self-Citations	1.831	1.473	-0.358 ***
	(2.998)	(2.507)	
Non-Self	9.217	7.65	-1.567 ***
	(24.590)	(16.423)	
From Articles with Female Last Author	2.062	2.263	0.201 ***
	(5.598)	(5.139)	
From Articles with Male Last Author	6.451	4.859	-1.592 ***
	(17.189)	(11.017)	
From Articles with Female First Author	3.113	3.156	0.042 ***
	(8.399)	(6.978)	
From Articles with Male First Author	5.255	3.855	-1.400 ***
	(14.049)	(8.815)	
Share of Articles Receiving No Citations, Three-Year Window	11.1%	13.1%	2.0% ***

(B)	First Author		
	Male	Female	Female-Male
Number of Articles	1,544,727	888,079	
	63.5%	36.5%	
Number of Forward Citations, Three-Year Window:			
All	10.93	9.871	-1.059 ***
	(23.601)	(24.469)	
Self-Citations	1.812	1.607	-0.205 ***
	(3.012)	(2.634)	
Non-Self	9.118	8.264	-0.854 ***
	(22.239)	(23.594)	
From Articles with Female Last Author	1.999	2.317	0.318 ***
	(5.022)	(6.197)	
From Articles with Male Last Author	6.421	5.36	-1.062 ***
	(15.998)	(15.490)	
From Articles with Female First Author	3.016	3.314	0.298 ***
	(7.583)	(8.803)	
From Articles with Male First Author	5.262	4.238	-1.024 ***
	(13.052)	(12.605)	
Share of Articles Receiving No Citations, Three-Year Window	11.8%	11.3%	-0.5% ***

Notes: There are on average 0.658 (7.6% of) citations for which we cannot identify the last author gender, and 0.793 (9.0% of) citations for which we cannot identify the first author gender. These rates of genderization are consistent with the genderization rate in our sample of cited articles. *** $p < 0.01$ from two-sided t-test. We classify cited and citing articles by *last* author gender for panel (A) and by *first* author gender for panel (B).

Table 3. OLS regression of three-year forward citations on author gender of cited article, observables and fixed effects

(A)	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with								
Dependent Variable	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Last Author	0.201***	-1.592***	-1.391***	0.134***	-0.418***	-0.284***	0.089***	-0.348***	-0.259***
	(0.035)	(0.133)	(0.164)	(0.016)	(0.042)	(0.053)	(0.015)	(0.040)	(0.053)
Author Team Size				0.107***	0.343***	0.450***	0.100***	0.317***	0.416***
				(0.013)	(0.047)	(0.060)	(0.012)	(0.044)	(0.056)
Number of References				0.028***	0.072***	0.100***	0.026***	0.069***	0.095***
				(0.001)	(0.004)	(0.006)	(0.001)	(0.004)	(0.006)
Number of Pages				0.055***	0.164***	0.219***	0.058***	0.169***	0.227***
				(0.006)	(0.021)	(0.027)	(0.007)	(0.021)	(0.028)
First Author Experience				0.007***	0.012***	0.019***	0.008***	0.016***	0.023***
				(0.001)	(0.004)	(0.005)	(0.001)	(0.004)	(0.005)
First Author Cumulated				-0.004*	-0.010	-0.014*	-0.004*	-0.010	-0.015*
Publications				(0.002)	(0.006)	(0.008)	(0.002)	(0.006)	(0.009)
First Author Cumulated				0.014***	0.037***	0.051***	0.014***	0.037***	0.051***
Citations/100				(0.004)	(0.013)	(0.017)	(0.005)	(0.013)	(0.018)
Last Author Experience				0.001	-0.010***	-0.009***	0.002*	-0.005***	-0.004
				(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
Last Author Cumulated				-0.004***	-0.011***	-0.015***	-0.004***	-0.010***	-0.015***
Publications				(0.001)	(0.002)	(0.003)	(0.001)	(0.003)	(0.004)
Last Author Cumulated				0.011***	0.029***	0.041***	0.011***	0.028***	0.038***
Citations/100				(0.002)	(0.005)	(0.007)	(0.002)	(0.005)	(0.007)
Journal Impact Factor				0.132***	0.416***	0.548***	0.134***	0.406***	0.540***
				(0.039)	(0.079)	(0.114)	(0.041)	(0.083)	(0.120)
Year FE	N	N	N	Y	Y	Y	Y	Y	Y
Journal FE	N	N	N	Y	Y	Y	Y	Y	Y
Affiliation FE	N	N	N	Y	Y	Y	Y	Y	Y
Field FE	N	N	N	Y	Y	Y	N	N	N
Concept FE	N	N	N	N	N	N	Y	Y	Y
Observations	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806
R-squared	0.000	0.002	0.001	0.191	0.224	0.221	0.223	0.263	0.258

(B)	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with								
Dependent Variable	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female First Author	0.298***	-1.024***	-0.726***	0.145***	-0.332***	-0.187***	0.105***	-0.278***	-0.173***
	(0.048)	(0.093)	(0.136)	(0.017)	(0.024)	(0.036)	(0.015)	(0.020)	(0.033)
Author Team Size				0.166***	0.278***	0.444***	0.153***	0.258***	0.462***
				(0.021)	(0.038)	(0.058)	(0.019)	(0.036)	(0.065)
Number of References				0.042***	0.057***	0.098***	0.040***	0.054***	0.103***
				(0.002)	(0.003)	(0.006)	(0.002)	(0.003)	(0.007)
Number of Pages				0.077***	0.140***	0.217***	0.082***	0.143***	0.210***
				(0.010)	(0.017)	(0.026)	(0.010)	(0.018)	(0.028)
First Author Experience				0.009***	0.009***	0.018***	0.011***	0.012***	0.015**
				(0.002)	(0.003)	(0.005)	(0.002)	(0.003)	(0.006)
First Author Cumulated Publications				-0.006*	-0.008	-0.014*	-0.006*	-0.008	-0.014*
				(0.003)	(0.005)	(0.008)	(0.003)	(0.005)	(0.008)
First Author Cumulated Citations/100				0.020***	0.029***	0.049***	0.020***	0.030***	0.050***
				(0.006)	(0.010)	(0.017)	(0.007)	(0.011)	(0.017)
Last Author Experience				-0.002	-0.007***	-0.009***	0.000	-0.004***	-0.008**
				(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.003)
Last Author Cumulated Publications				-0.007***	-0.008***	-0.015***	-0.006***	-0.008***	-0.015***
				(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.004)
Last Author Cumulated Citations/100				0.017***	0.023***	0.041***	0.016***	0.022***	0.042***
				(0.003)	(0.005)	(0.008)	(0.003)	(0.005)	(0.008)
Journal Impact Factor				0.204***	0.335***	0.539***	0.207***	0.326***	0.535***
				(0.055)	(0.065)	(0.116)	(0.057)	(0.069)	(0.116)
Year FE	N	N	N	Y	Y	Y	Y	Y	Y
Journal FE	N	N	N	Y	Y	Y	Y	Y	Y
Affiliation FE	N	N	N	Y	Y	Y	Y	Y	Y
Field FE	N	N	N	Y	Y	Y	N	N	N
Concept FE	N	N	N	N	N	N	Y	Y	Y
Observations	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806
R-squared	0.000	0.001	0.000	0.207	0.219	0.219	0.241	0.257	0.216

Notes: Robust standard errors clustered at journal level in parentheses: *** p<0.01, ** p<0.05, * p<0.1. We classify cited and citing articles by *last* author gender for panel (A) and by *first* author gender for panel (B)

Table 4. Top 10 (A) men's and (B) women's major fields in the life sciences

(A) Rank	Field	Num. of Articles	Share of Articles with Female Last Author	Share of Citations from Other Female: Written by Male Last Author (1)	Written by Female Last Author (2)	Gap (1)-(2)	P-Value
1	Optics	18,625	10.1%	13.0%	15.8%	-2.9%	0.00
				(0.24)	(0.26)		
2	Surgery	98,887	13.4%	14.7%	21.0%	-6.3%	0.00
				(0.24)	(0.28)		
3	Cardiology	33,799	13.5%	15.4%	21.9%	-6.5%	0.00
				(0.24)	(0.28)		
4	Stereochemistry	15,677	14.2%	17.7%	20.5%	-2.8%	0.00
				(0.25)	(0.27)		
5	Artificial intelligence	15,813	15.9%	18.0%	22.8%	-4.9%	0.00
				(0.25)	(0.28)		
6	Chemical engineering	23,273	17.2%	18.9%	21.1%	-2.2%	0.00
				(0.25)	(0.26)		
7	Biophysics	36,128	17.4%	19.9%	22.7%	-2.8%	0.00
				(0.22)	(0.24)		
8	Radiology	48,813	18.6%	18.4%	25.7%	-7.3%	0.00
				(0.26)	(0.29)		
9	Computational biology	36,776	18.9%	21.5%	24.8%	-3.3%	0.00
				(0.22)	(0.24)		
10	Chromatography	21,548	19.2%	22.0%	26.6%	-4.6%	0.00
				(0.27)	(0.29)		

(B) Rank	Field	Num. of Articles	Share of Articles with Female Last Author	Share of Citations from Other Female: Written by Male Last Author (1)	Written by Female Last Author (2)	Gap (1)-(2)	P-Value
1	Nursing	28,663	66.7%	52.4%	66.1%	-13.8%	0.00
				(0.35)	(0.35)		
2	Developmental psychology	25,395	53.2%	47.7%	53.9%	-6.1%	0.00
				(0.32)	(0.32)		
3	Gerontology	16,544	52.3%	45.5%	52.2%	-6.7%	0.00
				(0.31)	(0.32)		
4	Medical education	35,489	47.0%	39.2%	52.9%	-13.6%	0.00
				(0.35)	(0.37)		
5	Clinical psychology	56,911	45.4%	42.2%	48.8%	-6.6%	0.00
				(0.30)	(0.31)		
6	Family medicine	51,891	44.5%	41.4%	50.3%	-8.9%	0.00
				(0.32)	(0.33)		
7	Obstetrics	16,994	44.4%	39.6%	48.8%	-9.2%	0.00
				(0.31)	(0.32)		
8	Demography	24,821	42.7%	40.1%	46.7%	-6.6%	0.00
				(0.30)	(0.31)		
9	Environmental health	22,900	40.5%	37.4%	43.8%	-6.5%	0.00
				(0.29)	(0.31)		
10	Pediatrics	22,121	36.4%	34.4%	39.9%	-5.5%	0.00
				(0.30)	(0.31)		

Notes: We rank fields with more than 15,000 articles according to the share of articles with female last author, and report the bottom 10 in panel (A) and the top 10 in panel (B).

Table 5. Robustness Checks – using gender of authors at other positions to classify cited articles.

(A) Dependent Variable	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with							
	Female *Last* Author	Male *Last* Author	Female *Last* Author	Male *Last* Author	Female *Last* Author	Male *Last* Author	Female *Last* Author	Male *Last* Author
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Last Author Female	0.079***	-0.309***	0.069***	-0.296***				
	(0.015)	(0.040)	(0.015)	(0.040)				
N Female Authors in Non-Last Positions	0.064***	-0.239***						
	(0.007)	(0.023)						
First Author Female			0.086***	-0.170***				
			(0.010)	(0.019)				
N Female Authors in Middle Positions			0.055***	-0.216***				
			(0.006)	(0.024)				
Minority Female					0.044***	-0.022		
					(0.016)	(0.044)		
Majority Female					0.168***	-0.353***		
					(0.020)	(0.040)		
All Female					0.198***	-0.292***		
					(0.016)	(0.036)		
Share of Female							0.265***	-0.560***
							(0.025)	(0.047)
Control Variables	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y	Y	Y
Last Author Affiliation FE	Y	Y	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806
R-squared	0.223	0.263	0.223	0.263	0.223	0.263	0.223	0.263

(B) Dependent Variable	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with							
	Female *First* Author	Male *First* Author	Female *First* Author	Male *First* Author	Female *First* Author	Male *First* Author	Female *First* Author	Male *First* Author
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First Author Female	0.098***	-0.234***	0.085***	-0.178***				
	(0.014)	(0.017)	(0.014)	(0.016)				
N Female Authors in Non-First Positions	0.044***	-0.245***						
	(0.009)	(0.019)						
Last Author Female			0.045**	-0.276***				
			(0.021)	(0.033)				
N Female Authors in Middle Positions			0.047***	-0.212***				
			(0.009)	(0.021)				
Minority Female					0.058**	-0.038		
					(0.024)	(0.036)		
Majority Female					0.173***	-0.365***		
					(0.027)	(0.033)		
All Female					0.204***	-0.323***		
					(0.023)	(0.029)		
Share of Female							0.263***	-0.577***
							(0.032)	(0.039)
Control Variables	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806	2,432,806
R-squared	0.241	0.258	0.241	0.258	0.241	0.257	0.241	0.257

Notes: Robust standard errors clustered at journal level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We classify citing articles by *last* author for panel (A) and by *first* author gender for panel (B).

Table 6. Robustness Checks -- using gender of authors at other positions to classify both cited and citing articles.

Dependent Variable	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with			
	All Male	Minority Female	Majority Female	All Female
	(1)	(2)	(3)	(4)
All Male articles used as reference group				
Minority Female	-0.057***	0.010	0.051***	0.019***
	(0.016)	(0.023)	(0.018)	(0.005)
Majority Female	-0.258***	-0.166***	0.146***	0.091***
	(0.016)	(0.023)	(0.020)	(0.007)
All Female	-0.230***	-0.110***	0.120***	0.126***
	(0.014)	(0.020)	(0.017)	(0.006)
Control Variables	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y
Last Author Affiliation FE	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y
Observations	2,432,806	2,432,806	2,432,806	2,432,806
R-squared	0.270	0.202	0.221	0.175

Notes: Robust standard errors clustered at journal level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Gender differences in direct collaborators

(A)	Last Author Gender		Gap	
	Male	Female		
Composition of the collaborators of author team (Three-Year Window, 2005-2017)				
Total	141.655	117.944	-23.711	***
	(218.706)	(215.017)		
Number of Female	43.976	43.912	-0.064	(p=0.65)
	(74.253)	(81.663)		
Number of Male	86.375	65.737	-20.638	***
	(131.197)	(122.026)		
Number of Gender Unidentified	11.304	8.295	-3.009	***
	(25.508)	(20.544)		
Female Share	32.4%	42.6%	10.3%	***

(B)	First Author Gender		Gap	
	Male	Female		
Composition of the collaborators of author team (Three-Year Window, 2005-2017)				
Total	134.776	136.217	1.441	***
	(214.020)	(224.480)		
Number of Female	41.072	48.817	7.745	***
	(71.738)	(83.205)		
Number of Male	82.839	77.518	-5.321	***
	(129.242)	(128.877)		
Number of Gender Unidentified	10.865	9.882	-0.983	***
	(25.004)	(23.101)		
Female Share	31.8%	40.6%	8.80%	***

Notes: For each cited article, the direct collaborators include distinct researchers who have published a paper with one or more authors of the article's author team in the three years prior to the article's publication. The female share in the last row of each panel is calculated as the number of collaborators genderized as female divided by the number of all genderized collaborators. Standard deviation in parentheses. *** $p < 0.01$ from two-sided t-test. We classify cited articles by *last* author for panel (A) and by *first* author gender for panel (B).

Table 8. Gender Homophily and the Gender Composition of Direct Collaborators

(A) Dependent Variable	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with					
	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Female Last Author	0.092***	-0.339***	-0.247***	0.076***	-0.277***	-0.201***
	(0.017)	(0.044)	(0.059)	(0.017)	(0.042)	(0.057)
Number of Male Collaborators				-0.001***	0.012***	0.011***
				(0.000)	(0.001)	(0.001)
Number of Female Collaborators				0.004***	-0.013***	-0.009***
				(0.000)	(0.001)	(0.002)
Control Variables	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y
Affiliation FE	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y
Observations	2,063,054	2,063,054	2,063,054	2,063,054	2,063,054	2,063,054
R-squared	0.220	0.257	0.252	0.220	0.258	0.252

(B) Dependent Variable	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with					
	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Female First Author	0.109***	-0.267***	-0.158***	0.099***	-0.215***	-0.117***
	(0.018)	(0.021)	(0.035)	(0.017)	(0.020)	(0.034)
Number of Male Collaborators				0.000	0.010***	0.011***
				(0.000)	(0.001)	(0.001)
Number of Female Collaborators				0.003***	-0.012***	-0.009***
				(0.001)	(0.001)	(0.002)
Control Variables	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y
Affiliation FE	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y
Observations	2,063,883	2,063,883	2,063,883	2,063,883	2,063,883	2,063,883
R-squared	0.235	0.251	0.249	0.236	0.252	0.250

Notes: Robust standard errors clustered at journal level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We classify cited and citing articles by *last* author for panel (A) and by *first* author gender for panel (B).

Table 9. Gender Homophily in Forward Citations is *not* driven by discrimination based on gender inferred from first names

Sample:	Articles with ***last*** author who occasionally don't report full first names			Articles with ***first*** author who occasionally don't report full first names		
Dependent Variable	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with					
	Female Last Author	Male Last Author	Both	Female First Author	Male First Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Last Author Reporting Only Initials	0.031	-0.048	-0.016			
	(0.034)	(0.101)	(0.132)			
Last Author Female x Last Author Reporting Only Initials	-0.087	-0.019	-0.105			
	(0.053)	(0.116)	(0.162)			
First Author Reporting Only Initials				0.019	-0.043	-0.024
				(0.046)	(0.079)	(0.120)
First Author Female x First Author Reporting Only Initials				0.020	0.084	0.104
				(0.056)	(0.075)	(0.122)
Control Variables	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y
Affiliation FE	Y	Y	Y	Y	Y	Y
Last Author FE	Y	Y	Y	N	N	N
First Author FE	N	N	N	Y	Y	Y
Observations	796,544	796,544	796,544	381,187	381,187	381,187
R-squared	0.286	0.307	0.304	0.414	0.460	0.448

Notes: Robust standard errors clustered at journal level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We classify both cited and citing using *last* author gender in columns 1-3 and using *first* author gender in columns 4-6.

Table 10. Matched sample analysis based on the PubMed Related Citations Algorithm

(A) Dep. Var.	Number of Three-Year Forward Citations (Self-Citation Excluded), from					
	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Female Last Author	0.107***	-0.400***	-0.293***	0.007	-0.213***	-0.206***
	(0.031)	(0.054)	(0.081)	(0.024)	(0.049)	(0.070)
Matched Pair FE	N	N	N	Y	Y	Y
Control Variables	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Affiliation FE	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y
Observations	3,722,399	3,722,399	3,722,399	3,717,904	3,717,904	3,717,904
R-squared	0.244	0.305	0.294	0.669	0.699	0.693

(B) Dep. Var.	Number of Three-Year Forward Citations (Self-Citation Excluded), from					
	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Female First Author	0.123***	-0.314***	-0.191***	0.038*	-0.181***	-0.142***
	(0.025)	(0.025)	(0.043)	(0.021)	(0.024)	(0.041)
Matched Pair FE	N	N	N	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Affiliation FE	Y	Y	Y	Y	Y	Y
Concept FE	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y
Observations	3,722,702	3,722,702	3,722,702	3,718,516	3,718,516	3,718,516
R-squared	0.267	0.299	0.292	0.679	0.695	0.690

Notes: Robust standard errors clustered at journal level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We classify cited and citing articles by *last* author for panel (A) and by *first* author gender for panel (B).

Table A1. Data Cleaning Process & Attrition Rates

(A) Data Cleaning Process	Records Number	"Attrition" Rate	Notes
2002-2017 PubMed Records	13,359,358		
Restrict to Articles in English	11,369,223	85.1%	Other publication types such as reviews, letters, editorial materials, conference proceedings are excluded
Linkage to MAG	10,827,448	95.2%	Using the linkage provided by MAG
Winsorize Top 1% of Team Size	10,740,930	99.2%	The 99 percentile of the team size distribution is 17 authors. Large teams can have as many as 1074 authors.
With Journal Impact Factor, Num of Pages, Field ID missing.	**9,815,086**	91.4%	238,750 with Journal Impact Factor missing, 622,760 with Page Numbers missing, 103,298 with Field ID missing.

(B) Genderization of First and Last Author	First Author Country				Total	
	Non-US		US			
Both Gendered	5,545,308	79.3%	**2,454,886**	87.0%	8,000,194	81.5%
Only First Gendered	505,649	7.2%	118,441	4.2%	624,090	6.4%
Only Last Gendered	554,452	7.9%	190,813	6.8%	745,265	7.6%
Neither Gendered	386,952	5.5%	58,585	2.1%	445,537	4.5%
Total	6,992,361	100.0%	2,822,725	100.0%	**9,815,086**	100.0%

Notes: In the end, when we run regressions with full controls, 22,080 singleton observations are dropped due to our extensive controls of fixed effects, leaving us with a **final sample of 2,432,806 articles**.

Table A2. Split sample analysis by concept-level female share of (A) *last* authors and (B) *first* authors.

(A)	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with		
Dependent Variable	Female Last Author	Male Last Author	Both
Sample: Concepts with female share of last authors <10%			
	(1)	(2)	(3)
Female Last Author	-0.040	-0.423***	-0.463***
	(0.025)	(0.098)	(0.117)
Observations	198,635	198,635	198,635
R-squared	0.189	0.215	0.213
Sample: Concepts with female share of last authors [10%,20%)			
	(4)	(5)	(6)
Female Last Author	0.021	-0.448***	-0.427***
	(0.014)	(0.052)	(0.064)
Observations	619,786	619,786	619,786
R-squared	0.228	0.253	0.255
Sample: Concepts with female share of last authors [20%,30%)			
	(7)	(8)	(9)
Female Last Author	0.050**	-0.451***	-0.402***
	(0.020)	(0.062)	(0.079)
Observations	851,310	851,310	851,310
R-squared	0.229	0.253	0.254
Sample: Concepts with female share of last authors [30%,40%)			
	(10)	(11)	(12)
Female Last Author	0.093***	-0.291***	-0.198***
	(0.020)	(0.050)	(0.068)
Observations	364,294	364,294	364,294
R-squared	0.191	0.197	0.200
Sample: Concepts with female share of last authors [40%,50%)			
	(13)	(14)	(15)
Female Last Author	0.143***	-0.211***	-0.069
	(0.036)	(0.046)	(0.079)
Observations	227,263	227,263	227,263
R-squared	0.176	0.218	0.207
Sample: Concepts with female share of last authors >=50%			
	(16)	(17)	(18)
Female Last Author	0.151***	-0.052	0.100
	(0.046)	(0.055)	(0.097)
Observations	153,523	153,523	153,523
R-squared	0.144	0.256	0.205

(B) Dependent Variable	Number of Three-Year Forward Citations (Self-Citation Excluded), From Articles with Female First Author	Male First Author	Both
Sample: Concepts with female share of first authors <10%			
	(1)	(2)	(3)
Female First Author	-0.181**	-0.410***	-0.591***
	(0.071)	(0.151)	(0.210)
Observations	63,412	63,412	63,412
R-squared	0.276	0.249	0.262
Sample: Concepts with female share of first authors [10%,20%)			
	(4)	(5)	(6)
Female First Author	0.031	-0.137**	-0.105
	(0.031)	(0.060)	(0.087)
Observations	201,409	201,409	201,409
R-squared	0.172	0.208	0.201
Sample: Concepts with female share of first authors [20%,30%)			
	(7)	(8)	(9)
Female First Author	0.007	-0.311***	-0.304***
	(0.021)	(0.042)	(0.061)
Observations	479,797	479,797	479,797
R-squared	0.187	0.214	0.209
Sample: Concepts with female share of first authors [30%,40%)			
	(10)	(11)	(12)
Female First Author	0.025	-0.333***	-0.307***
	(0.019)	(0.033)	(0.050)
Observations	799,057	799,057	799,057
R-squared	0.274	0.286	0.288
Sample: Concepts with female share of first authors [40%,50%)			
	(13)	(14)	(15)
Female First Author	0.111***	-0.252***	-0.141**
	(0.025)	(0.035)	(0.057)
Observations	473,129	473,129	473,129
R-squared	0.230	0.238	0.242
Sample: Concepts with female share of first authors >=50%			
	(16)	(17)	(18)
Female First Author	0.229***	-0.223***	0.006
	(0.052)	(0.039)	(0.088)
Observations	400,831	400,831	400,831
R-squared	0.164	0.185	0.181

Notes: We use the MAG's scientific concept with the highest score as the one that best describes the research focus of an article. Robust standard errors clustered at journal level in parentheses *** p<0.01, ** p<0.05, * p<0.1; all specifications include field fixed effects, year fixed effects, journal fixed effects, and first author's affiliation fixed effects, as well as author team size, number of references, number of pages, journal impact factor, last author and first authors' experience in years, last and first authors' cumulated publications, and last and first authors' cumulated citations as controls. We classify cited and citing articles by *last* author for panel (A) and by *first* author gender for panel (B).

Table A3. Split sample analysis by authors' entry cohorts.

(A)	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with		
Dependent Variable	Female Last Author	Male Last Author	Both
Sample: Entry In 1975-1984	(1)	(2)	(3)
Female Last Author	0.233***	-0.512***	-0.279**
	(0.056)	(0.082)	(0.129)
Observations	586,326	586,326	586,326
R-squared	0.205	0.273	0.266
Sample: Entry In 1985-1989	(4)	(5)	(6)
Female Last Author	0.133***	-0.459***	-0.326***
	(0.026)	(0.075)	(0.095)
Observations	308,748	308,748	308,748
R-squared	0.269	0.281	0.285
Sample: Entry In 1990-1994	(7)	(8)	(9)
Female Last Author	0.128***	-0.567***	-0.439***
	(0.026)	(0.092)	(0.114)
Observations	337,779	337,779	337,779
R-squared	0.218	0.238	0.238
Sample: Entry In 1995-1999	(10)	(11)	(12)
Female Last Author	0.159***	-0.389***	-0.230***
	(0.019)	(0.048)	(0.059)
Observations	348,638	348,638	348,638
R-squared	0.227	0.262	0.261
Sample: Entry In 2000-2004	(13)	(14)	(15)
Female Last Author	0.104***	-0.329***	-0.225***
	(0.018)	(0.049)	(0.064)
Observations	383,996	383,996	383,996
R-squared	0.232	0.227	0.230
Sample: Entry In 2005-2009	(16)	(17)	(18)
Female Last Author	0.138***	-0.203***	-0.065
	(0.016)	(0.040)	(0.053)
Observations	258,406	258,406	258,406
R-squared	0.224	0.243	0.243
Sample: Entry In 2010-2014	(19)	(20)	(21)
Female Last Author	0.129***	-0.155***	-0.026
	(0.020)	(0.040)	(0.056)
Observations	142,758	142,758	142,758
R-squared	0.250	0.265	0.268
Sample: Entry In 2015-2017	(22)	(23)	(24)
Female Last Author	0.112***	-0.126*	-0.013
	(0.036)	(0.064)	(0.094)
Observations	44,923	44,923	44,923
R-squared	0.329	0.364	0.362

(B)	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with		
Dependent Variable	Female First Author	Male First Author	Both
Sample: Entry In 1975-1984	(1)	(2)	(3)
Female First Author	0.264**	-0.632***	-0.368**
	(0.103)	(0.094)	(0.187)
Observations	198,729	198,729	198,729
R-squared	0.236	0.266	0.262
Sample: Entry In 1985-1989	(4)	(5)	(6)
Female First Author	0.183***	-0.561***	-0.378***
	(0.060)	(0.094)	(0.144)
Observations	114,111	114,111	114,111
R-squared	0.213	0.215	0.217
Sample: Entry In 1990-1994	(7)	(8)	(9)
Female First Author	0.032	-0.754***	-0.723***
	(0.049)	(0.096)	(0.137)
Observations	155,441	155,441	155,441
R-squared	0.255	0.255	0.262
Sample: Entry In 1995-1999	(10)	(11)	(12)
Female First Author	0.124**	-0.502***	-0.378***
	(0.050)	(0.058)	(0.097)
Observations	248,852	248,852	248,852
R-squared	0.242	0.290	0.281
Sample: Entry In 2000-2004	(13)	(14)	(15)
Female First Author	0.126***	-0.321***	-0.195***
	(0.027)	(0.044)	(0.067)
Observations	522,165	522,165	522,165
R-squared	0.247	0.270	0.271
Sample: Entry In 2005-2009	(16)	(17)	(18)
Female First Author	0.113***	-0.284***	-0.171***
	(0.023)	(0.036)	(0.056)
Observations	587,160	587,160	587,160
R-squared	0.305	0.294	0.304
Sample: Entry In 2010-2014	(19)	(20)	(21)
Female First Author	0.197***	-0.133***	0.064
	(0.022)	(0.034)	(0.053)
Observations	449,922	449,922	449,922
R-squared	0.224	0.241	0.239
Sample: Entry In 2015-2017	(22)	(23)	(24)
Female First Author	0.165***	-0.088***	0.078
	(0.025)	(0.032)	(0.053)
Observations	139,670	139,670	139,670
R-squared	0.275	0.258	0.275

Notes: We define an author's entry cohort according to the author's first year of publication in the matched dataset between PubMed and MAG, which starts from 1975. Robust standard errors clustered at journal level in parentheses *** p<0.01, ** p<0.05, * p<0.1. All specifications include field fixed effects, year fixed effects, journal fixed effects, and first author's affiliation fixed effects, as well as author team size, number of references, number of pages, journal impact factor, last author and first authors' experience in years, last and first authors' cumulated publications, and last and first authors' cumulated citations as controls. We classify cited and citing article by *last* author for panel (A) and by *first* author gender for panel (B).

Table A4. Robustness – Poisson Model Specifications

Poisson Regressions	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with					
Dependent Variable:	Female Last Author	Male Last Author	Both	Female First Author	Male First Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Female Last Author	0.049***	-0.048***	-0.020***			
	(0.004)	(0.004)	(0.004)			
Female Frist Author				0.036***	-0.050***	-0.015***
				(0.004)	(0.003)	(0.003)
Author Team Size	0.030***	0.032***	0.032***	0.030***	0.033***	0.032***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Number of References	0.004***	0.003***	0.004***	0.004***	0.003***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of Pages	0.032***	0.033***	0.032***	0.032***	0.033***	0.032***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
First Author Experience	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First Author Cumulated Publications	-0.000	0.000*	0.000	0.000	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First Author Cumulated Citations/100	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Last Author Experience	0.001**	-0.001**	-0.000	0.000	-0.001**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Last Author Cumulated Publications	-0.000***	-0.000**	-0.000***	-0.000***	-0.000**	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Last Author Cumulated Citations/100	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Journal Impact Factor	0.006	0.009	0.008	0.007	0.008	0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
Year FE	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y
Affiliation FE	Y	Y	Y	Y	Y	Y
Field FE	Y	Y	Y	Y	Y	Y
Observations	2,402,568	2,421,777	2,424,712	2,410,832	2,421,197	2,425,571

Notes: Robust standard errors clustered at journal level in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table A5. Robustness – Additional controls for middle authors' characteristics

(A)	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with								
Dep. Var.:	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Last Author	0.077***	-0.396***	-0.319***	0.075***	-0.405***	-0.330***	0.077***	-0.397***	-0.321***
	(0.020)	(0.053)	(0.070)	(0.020)	(0.054)	(0.071)	(0.020)	(0.053)	(0.070)
Middle Author Characteristics									
Share of Middle Authors							0.126***	0.325***	0.451***
from the Top 100 Affiliations							(0.028)	(0.079)	(0.104)
Average Years of Experience									
Average Cumulated Publications									
Average Cumulated Citations in 100									
Max. of Years of Experience									
Max. of Cumulated Publications									
Max. of Cumulated Citations in 100									
Last Author's Aff. FE	Y	Y	Y	\	\	\	Y	Y	Y
FE. Of Best Aff. of Among Mid-Authors	\	\	\	Y	Y	Y	\	\	\
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1,745,325	1,745,325	1,745,325	1,746,377	1,746,377	1,746,377	1,745,325	1,745,325	1,745,325
R-squared	0.234	0.277	0.271	0.232	0.274	0.269	0.234	0.277	0.271

(A) Continued	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with					
	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both
	(10)	(11)	(12)	(13)	(14)	(15)
Female Last Author	0.073***	-0.405***	-0.332***	0.074***	-0.403***	-0.329***
	(0.019)	(0.051)	(0.068)	(0.019)	(0.051)	(0.067)
Middle Author Characteristics						
Share of Middle Authors	0.111***	0.288***	0.398***	0.113***	0.298***	0.411***
from the Top 100 Affiliations	(0.027)	(0.074)	(0.098)	(0.026)	(0.074)	(0.097)
Average Years of Experience	0.002	-0.003	-0.001			
	(0.002)	(0.005)	(0.007)			
Average Cumulated Publications	-0.003***	-0.007***	-0.010***			
	(0.001)	(0.002)	(0.003)			
Average Cumulated Citations in 100	0.008***	0.019***	0.027***			
	(0.002)	(0.005)	(0.006)			
Max. of Years of Experience				-0.001	-0.009***	-0.010***
				(0.001)	(0.002)	(0.003)
Max. of Cumulated Publications				-0.001**	-0.002**	-0.003**
				(0.000)	(0.001)	(0.002)
Max. of Cumulated Citations in 100				0.003***	0.006***	0.009***
				(0.001)	(0.002)	(0.003)
Last Author's Aff. FE	Y	Y	Y	Y	Y	Y
FE. Of Best Aff. of Among Mid-Authors	\	\	\	\	\	\
Other Controls	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y
Field FE	Y	Y	Y	Y	Y	Y
Observations	1,745,202	1,745,202	1,745,202	1,745,202	1,745,202	1,745,202
R-squared	0.234	0.277	0.271	0.234	0.277	0.271

(B)	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with								
Dep. Var.:	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female First Author	0.103***	-0.285***	-0.182***	0.104***	-0.287***	-0.182***	0.102***	-0.285***	-0.183***
	(0.015)	(0.018)	(0.028)	(0.015)	(0.018)	(0.028)	(0.015)	(0.018)	(0.028)
Middle Author Characteristics							0.085**	0.112*	0.197**
Share of Middle Authors from the Top 100 Affiliations							(0.036)	(0.057)	(0.090)
Average Years of Experience									
Average Cumulated Publications									
Average Cumulated Citations in 100									
Max. of Years of Experience									
Max. of Cumulated Publications									
Max. of Cumulated Citations in 100									
First Author's Aff. FE									
FE. Of Best Aff. of Among Mid-Authors									
Other Controls									
Year FE									
Journal FE									
Field FE									
Observations	1,746,106	1,746,106	1,746,106	1,746,377	1,746,377	1,746,377	1,746,106	1,746,106	1,746,106
R-squared	0.251	0.272	0.269	0.249	0.270	0.267	0.251	0.272	0.269

(B) Continued	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with					
	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both
	(10)	(11)	(12)	(13)	(14)	(15)
Female First Author	0.101***	-0.287***	-0.186***	0.101***	-0.285***	-0.184***
	(0.015)	(0.018)	(0.027)	(0.015)	(0.018)	(0.027)
Middle Author Characteristics	0.069**	0.090	0.159*	0.069**	0.095*	0.165*
Share of Middle Authors from	(0.034)	(0.055)	(0.085)	(0.034)	(0.054)	(0.085)
the Top 100 Affiliations	0.001	-0.002	-0.000			
Average Years of Experience	(0.003)	(0.004)	(0.007)			
	-0.004***	-0.005***	-0.010***			
Average Cumulated Publications	(0.001)	(0.002)	(0.003)			
	0.011***	0.014***	0.025***			
Average Cumulated Citations in 100	(0.003)	(0.004)	(0.006)			
				-0.003*	-0.007***	-0.010***
Max. of Years of Experience				(0.001)	(0.002)	(0.003)
				-0.001**	-0.002*	-0.003**
Max. of Cumulated Publications				(0.001)	(0.001)	(0.002)
				0.004***	0.004**	0.008***
Max. of Cumulated Citations in 100				(0.001)	(0.002)	(0.003)
First Author's Aff. FE						
FE. Of Best Aff. of Among Mid-Authors						
Other Controls						
Year FE						
Journal FE						
Field FE						
Observations	1,745,983	1,745,983	1,745,983	1,745,983	1,745,983	1,745,983
R-squared	0.251	0.272	0.269	0.251	0.272	0.269

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. The sample used for this table is restricted to articles with at least one middle author, i.e. articles with at least three authors.

Table A6. Robustness – Expanded sample specification including OECD countries minus South Korea and Japan.

(A)	Dep. Var.	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with		
		Female Last Author	Male Last Author	Both
Sample A: Japan, Korea, China		(1)	(2)	(3)
Observations: 770,686	Female Last Author	0.054***	-0.028	0.026
		(0.013)	(0.035)	(0.047)
	R-squared	0.262	0.280	0.281
Sample B: OECD (Excluding US, Japan, Korea)		(4)	(5)	(6)
Observations: 2,780,960	Female Last Author	0.087***	-0.209***	-0.122***
		(0.007)	(0.018)	(0.023)
	R-squared	0.281	0.331	0.328
Sample C: Rest of the World		(7)	(8)	(9)
Observations: 1,925,130	Female Last Author	0.025***	-0.241***	-0.216***
		(0.008)	(0.026)	(0.033)
	R-squared	0.223	0.250	0.250
Sample D: US		(10)	(11)	(12)
Observations: 2,432,806	Female Last Author	0.089***	-0.348***	-0.259***
		(0.015)	(0.040)	(0.053)
	R-squared	0.223	0.263	0.258
	Other Controls	Y	Y	Y
	Year FE	Y	Y	Y
	Journal FE	Y	Y	Y
	Concept FE	Y	Y	Y
	Last Author's Aff. FE	Y	Y	Y

(B)	Dep. Var.	Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with		
		Female First Author	Male First Author	Both
Sample A: Japan, Korea, China		(1)	(2)	(3)
Observations: 770,686	Female First Author	0.006	-0.123***	-0.118***
		(0.012)	(0.018)	(0.029)
	R-squared	0.265	0.270	0.273
Sample B: OECD (Excluding US, Japan, Korea)		(4)	(5)	(6)
Observations: 2,780,960	Female First Author	0.072***	-0.229***	-0.157***
		(0.009)	(0.013)	(0.019)
	R-squared	0.296	0.330	0.326
Sample C: Rest of the World		(7)	(8)	(9)
Observations: 1,925,130	Female First Author	0.034***	-0.236***	-0.202***
		(0.011)	(0.019)	(0.029)
	R-squared	0.229	0.239	0.241
Sample D: US		(10)	(11)	(12)
Observations: 2,432,806	Female First Author	0.105***	-0.278***	-0.173***
		(0.015)	(0.020)	(0.030)
	R-squared	0.241	0.257	0.256
	Other Controls	Y	Y	Y
	Year FE	Y	Y	Y
	Journal FE	Y	Y	Y
	Concept FE	Y	Y	Y
	First Author's Aff. FE	Y	Y	Y

Notes: Robust standard errors clustered at the journal level in parentheses, *** p<0.01, ** p<0.05, * p<0.1. We classify citing and cited articles by *last* author gender for panel (A) and by *first* author gender for panel (B).

Table A7. Robustness – Alternative Gender Imputation Criteria

(A) Criteria for Gender Imputation:	Baseline Frequency≥10			Laxer No restriction			Stricter Frequency≥10, Prob≥0.9		
Dependent Variable: Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both	Female Last Author	Male Last Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Last Author	0.132***	-0.481***	-0.349***	0.138***	-0.456***	-0.318***	0.135***	-0.418***	-0.284***
	(0.020)	(0.053)	(0.069)	(0.019)	(0.048)	(0.063)	(0.015)	(0.041)	(0.051)
Author Team Size	0.117***	0.369***	0.486***	0.112***	0.357***	0.469***	0.096***	0.327***	0.424***
	(0.014)	(0.049)	(0.063)	(0.013)	(0.047)	(0.061)	(0.011)	(0.043)	(0.055)
Number of References	0.029***	0.077***	0.106***	0.028***	0.073***	0.101***	0.024***	0.065***	0.089***
	(0.002)	(0.005)	(0.007)	(0.001)	(0.005)	(0.006)	(0.001)	(0.004)	(0.005)
Number of Pages	0.060***	0.178***	0.237***	0.057***	0.171***	0.228***	0.048***	0.155***	0.204***
	(0.008)	(0.024)	(0.032)	(0.007)	(0.023)	(0.030)	(0.006)	(0.020)	(0.025)
First Author Experience	0.007***	0.014***	0.021***	0.007***	0.014***	0.020***	0.006***	0.013***	0.019***
	(0.002)	(0.005)	(0.006)	(0.002)	(0.004)	(0.006)	(0.001)	(0.004)	(0.005)
First Author Cumulated Publications	-0.005*	-0.012	-0.017	-0.005*	-0.011	-0.015	-0.003**	-0.008	-0.012*
	(0.003)	(0.008)	(0.010)	(0.002)	(0.007)	(0.009)	(0.002)	(0.005)	(0.007)
First Author Cumulated Citations/100	0.015***	0.040***	0.055***	0.014***	0.038***	0.052***	0.011***	0.032***	0.043***
	(0.005)	(0.015)	(0.021)	(0.005)	(0.014)	(0.018)	(0.003)	(0.010)	(0.014)
Last Author Experience	0.001	-0.010***	-0.009***	0.001	-0.009***	-0.008***	0.001	-0.008***	-0.007***
	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
Last Author Cumulated Publications	-0.005***	-0.012***	-0.017***	-0.005***	-0.011***	-0.016***	-0.004***	-0.009***	-0.013***
	(0.001)	(0.003)	(0.004)	(0.001)	(0.003)	(0.004)	(0.001)	(0.002)	(0.003)
Last Author Cumulated Citations/100	0.012***	0.032***	0.044***	0.011***	0.030***	0.041***	0.009***	0.025***	0.034***
	(0.003)	(0.006)	(0.009)	(0.002)	(0.006)	(0.008)	(0.002)	(0.004)	(0.005)
Journal Impact Factor	0.163***	0.480***	0.643***	0.143***	0.422***	0.566***	0.098***	0.315***	0.413***
	(0.055)	(0.111)	(0.163)	(0.045)	(0.090)	(0.133)	(0.028)	(0.054)	(0.077)
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Affiliation FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1,924,091	1,924,091	1,924,091	1,924,091	1,924,091	1,924,091	1,924,091	1,924,091	1,924,091
R-squared	0.183	0.213	0.210	0.186	0.217	0.214	0.189	0.223	0.221

(B) Criteria for Gender Imputation:	Baseline Frequency≥10			Laxer No restriction			Stricter Frequency≥10, Prob≥0.9		
Dependent Variable: Number of Three-Year Forward Citations (Self-Citation Excluded), from Articles with	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both	Female First Author	Male First Author	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Frist Author	0.148***	-0.400***	-0.251***	0.150***	-0.382***	-0.232***	0.139***	-0.343***	-0.204***
	(0.022)	(0.030)	(0.045)	(0.021)	(0.027)	(0.041)	(0.017)	(0.023)	(0.033)
Author Team Size	0.180***	0.305***	0.485***	0.173***	0.291***	0.464***	0.147***	0.256***	0.402***
	(0.022)	(0.041)	(0.063)	(0.021)	(0.039)	(0.060)	(0.018)	(0.033)	(0.051)
Number of References	0.044***	0.061***	0.105***	0.042***	0.057***	0.099***	0.035***	0.048***	0.083***
	(0.002)	(0.004)	(0.007)	(0.002)	(0.004)	(0.006)	(0.002)	(0.003)	(0.005)
Number of Pages	0.084***	0.154***	0.237***	0.080***	0.146***	0.226***	0.066***	0.128***	0.194***
	(0.012)	(0.021)	(0.032)	(0.011)	(0.019)	(0.030)	(0.008)	(0.015)	(0.023)
First Author Experience	0.010***	0.010***	0.020***	0.010***	0.010***	0.019***	0.008***	0.010***	0.018***
	(0.002)	(0.004)	(0.006)	(0.002)	(0.003)	(0.006)	(0.002)	(0.003)	(0.004)
First Author Cumulated Publications	-0.007*	-0.009	-0.016	-0.006*	-0.008	-0.015*	-0.004*	-0.006	-0.010*
	(0.004)	(0.006)	(0.010)	(0.003)	(0.005)	(0.009)	(0.002)	(0.004)	(0.006)
First Author Cumulated Citations/100	0.021***	0.033***	0.054***	0.020***	0.030***	0.050***	0.016***	0.024***	0.039***
	(0.007)	(0.012)	(0.020)	(0.007)	(0.011)	(0.017)	(0.005)	(0.007)	(0.012)
Last Author Experience	-0.001	-0.007***	-0.008***	-0.001	-0.006***	-0.007***	-0.001	-0.005***	-0.006***
	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)
Last Author Cumulated Publications	-0.008***	-0.010***	-0.017***	-0.007***	-0.009***	-0.016***	-0.005***	-0.006***	-0.012***
	(0.002)	(0.003)	(0.005)	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)	(0.003)
	0.019***	0.026***	0.045***	0.018***	0.024***	0.041***	0.014***	0.018***	0.032***
Last Author Cumulated Citations/100	(0.004)	(0.006)	(0.010)	(0.004)	(0.005)	(0.009)	(0.002)	(0.003)	(0.006)
Journal Impact Factor	0.242***	0.395***	0.637***	0.218***	0.341***	0.559***	0.144***	0.221***	0.365***
	(0.074)	(0.096)	(0.167)	(0.066)	(0.077)	(0.139)	(0.039)	(0.042)	(0.075)
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Journal FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Affiliation FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	1,924,792	1,924,792	1,924,792	1,924,792	1,924,792	1,924,792	1,924,792	1,924,792	1,924,792
R-squared	0.198	0.208	0.208	0.199	0.211	0.212	0.205	0.218	0.220

Notes: The frequency used as criteria for gender imputation indicates the number of individuals with that given name with a known, validated and confirmed gender by Genderize.io in its database. Robust standard errors clustered at journal level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample is restricted to the observations where the *first* author's gender is identified using all the three kinds of gender imputation criteria to assure that the variations in coefficients are not due to sample variations. In each specification, the genders of cited and citing articles are identified using the same gender imputation thresholds. We classify cited and citing article by *last* author for panel (A) and by *first* author gender for panel (B).