RESEARCH ARTICLE

Female-led teams produce more innovative ideas yet receive less scientific impact

Alex J. Yang[1], Ying Ding[2], and Meijun Liu[3]

[1]School of Information Management, Nanjing University, Nanjing, China
[2]School of Information, The University of Texas at Austin, Austin, TX, USA
[3]Institute for Global Public Policy, Fudan University, Shanghai, China

Keywords: disruption, gender bias, novelty, science of science, scientific impact

ABSTRACT

Despite long-standing concerns about gender bias in science, there remains a lack of understanding regarding the performance of female scientists as team leaders compared to their male counterparts. This study explores differences between female-led and male-led teams in terms of scientific impact, novelty, and disruption, utilizing a comprehensive data set of journal articles spanning from 1980 to 2016 across STEM fields. We employed Coarsened Exact Matching (CEM) to match female and male scientists based on their characteristics. Subsequently, we applied multivariable logistic regression models to compare the outcomes of journal articles produced by female-led and male-led teams. Our analysis reveals that female-led teams generate more novel and disruptive ideas. However, they tend to produce articles with less scientific impact compared to their male-led counterparts. This suggests a systemic undervaluation of the contributions of female scientists. Further analysis indicates that this gender bias intensifies in later career stages and with larger team sizes. Additionally, significant field-specific heterogeneity is observed, with the most pronounced bias found in Biology and Medicine. These findings highlight the urgent need for policy adjustments to address these biases and promote a more equitable evaluation system in scientific research.

1. INTRODUCTION

The persistent underrepresentation of women in scientific leadership positions has been a topic of fervent scholarly discourse (Dion, Sumner, & Mitchell, 2018). Since 1901, only 26 of the 646 Nobel Prizes in Physics, Chemistry, and Medicine have been awarded to women[1]. The National Institutes of Health (NIH) Gender Inequality Task Force Report (2018)[2] also exposes concerning trends: Women comprise a mere 39% of tenure-track faculty positions and a meager 23% of tenured faculty roles, with leadership positions even more scarce. However, the underrepresentation of women cannot be explained by an insufficient pool of highly qualified women, because women have exceeded half of Ph.D. graduates in the biological sciences for more than 10 years[3].

For women who persevere in scientific careers, formidable challenges abound. Compelling evidence points towards gender bias in crucial aspects of academic life, including recruitment

[1] The Nobel laureate data is available at https://www.nobelprize.org/.
[2] The National Institutes of Health (NIH) Gender Inequality Task Force Report is available at https://diversity.nih.gov/general-page/national-institutes-health-nih-gender-inequality-task-force-report.
[3] See annual report of Council of Graduate Schools (CGS) at https://cgsnet.org/.

an open access journal

Citation: Yang, A. J., Ding, Y., & Liu, M. (2024). Female-led teams produce more innovative ideas yet receive less scientific impact. *Quantitative Science Studies*, *5*(4), 861–881. https://doi.org/10.1162/qss_a_00335

DOI: https://doi.org/10.1162/qss_a_00335

Peer Review: https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00335

Supporting Information: https://doi.org/10.1162/qss_a_00335

Received: 14 June 2024
Accepted: 10 October 2024

Corresponding Author:
Alex J. Yang
alexjieyang@outlook.com

Handling Editor:
Vincent Larivière

Copyright: © 2024 Alex J. Yang, Ying Ding, and Meijun Liu. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

practices (Moss-Racusin, Dovidio et al., 2012), salary structures (Ginther, 2003), access to research funding (Witteman, Hendricks et al., 2019), migration (Zhao, Akbaritabar et al., 2023), research strategies (Liu, Yang et al., 2023), and recognition through prestigious awards (Ma, Oliveira et al., 2019). Furthermore, women are demonstrably underrepresented in overall scientific output (Kong, Martin-Gutierrez, & Karimi, 2022), particularly in coveted first-authorship positions (Ductor, Goyal, & Prummer, 2023; Liu, Zhang et al., 2022; Ross, Glennon et al., 2022). The peer-review process itself is not immune to bias, as women receive unequal treatment in this process (Witteman et al., 2019). This phenomenon, aptly termed the "Matthew Matilda effect" (Rossiter, 1993), posits that established male researchers, dominating the publication landscape, garner a disproportionate share of citations, reputation, and influence within their fields. These disparities point towards a pervasive culture and climate that are unsupportive of women in science, coupled with ingrained structural issues hindering their career advancement. As an underrepresented group within STEM fields, women are confronting significant barriers.

Historically, the publication of research papers within STEM fields has exhibited a marked gender disparity, with men typically occupying a leading authorship position (Ross et al., 2022). However, over time, this disparity has shown signs of narrowing across all STEM disciplines (Figure 1). A significant increase is evident in the proportion of female scientists listed as the first author of research publications, rising from 15.6% in 1990 to 21.6% in 2016. A similar trend is observed for female scientists listed as the last author, with an increase from 10.1% to 21.2% during the same period. Furthermore, the distribution of female scientists across research teams and affiliations demonstrates a high degree of heterogeneity. Notably, a greater rate of increase in the proportion of female scientists as leading authors is observed within large research teams and highly ranked institutions.

The underrepresentation and undervaluation of female scientists in STEM fields have been long-standing issues, raising concerns about gender bias and its impact on scientific advancement (Lerman, Yu et al., 2022; Ross et al., 2022; Teich, Kim et al., 2022; Yang, Tian et al., 2022).

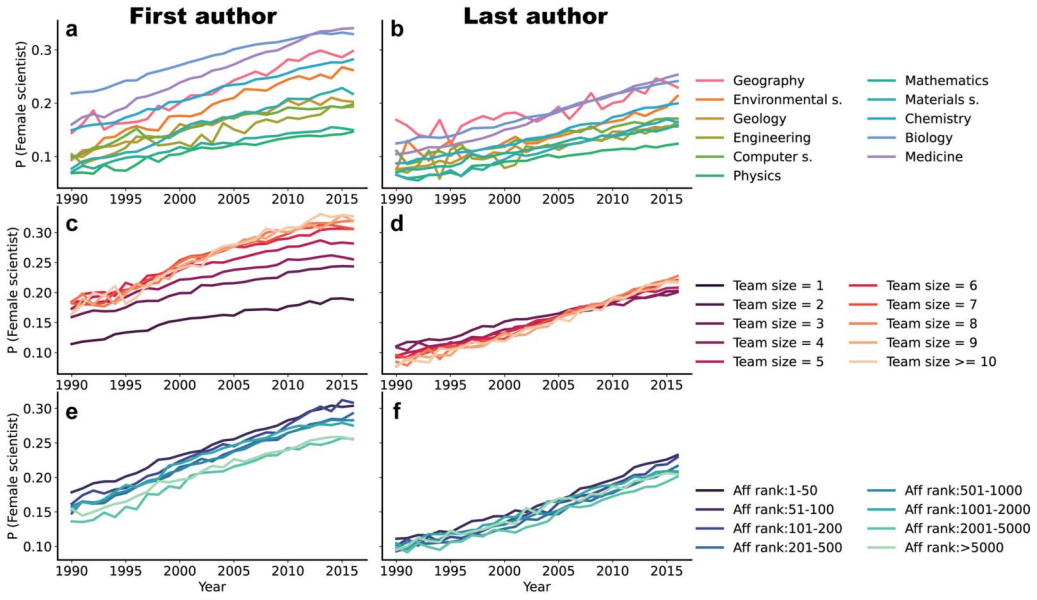


Figure 1. Increasing proportion of female-led teams in STEM fields. The proportion of female leaders, either as the first author or as the last author, shows an increasing trend across several categories: (a, b) different fields of study; (c, d) various team sizes; and (e, f) diverse affiliations. Notably, the proportion of female leaders is higher in (a, b) Biology and Medicine fields; (c, d) large teams; and (e, f) prestigious affiliations.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

While existing studies have explored gender differences in scientific productivity, citations, and recognition, they do not systematically analyze the differences in scientific impact and innovation when female scientists serve as team leaders. In most STEM fields, the first author is typically the individual who has made the most significant contribution to the research and has played a major role in the execution and writing of the study. The last author, often known as the senior or corresponding author, is usually the team leader who conceptualized the study, secured funding, and provided overall supervision (Shi, Liu, & Wang, 2023). By focusing on the gender of the first and last authors, we can effectively identify instances where female scientists are leading the research efforts either through direct contribution (first author) or by providing overarching leadership and guidance (last author).

Our study addresses these gaps by focusing on female scientists' performance as team leaders, comparing it to their male counterparts. We use a state-of-the-art methodological combining Coarsened Exact Matching (CEM) (Iacus, King, & Porro, 2012) and multivariable logistic regression, to ensure a robust and fair comparison. Additionally, we employ advanced measures of novelty (Uzzi, Mukherjee et al., 2013) and disruption (Funk & Owen-Smith, 2017; Park, Leahey, & Funk, 2023) based on complex network analysis, offering a more nuanced understanding of innovation in scientific research. These measures help us capture the extent to which female-led teams contribute to paradigm shifts and generate novel ideas, providing insights beyond traditional citation metrics.

This paper presents a rigorous investigation into the comparison of female-led teams and male-led teams within STEM fields between 1980 and 2016. Our inquiry delves into potential gender differences across three key metrics: scientific impact, atypical combinations of knowledge, and the potential to disrupt prevailing scientific paradigms. Our findings unveil a critical disparity: While female-led teams demonstrate a propensity for generating more novel ideas and challenging established paradigms, their work receives fewer citations compared to their male counterparts. This potentially reflects a systemic bias that undervalues the contributions of female scientists within the scientific community.

2. THEORETICAL BACKGROUND

2.1. Novel and Disruptive Ideas

Innovation drives scientific progress, propelling science into uncharted territories and expanding humanity's understanding of the natural and social world. In this paper, we consider two types of innovation in science: (ex-ante) novelty and (ex-post) disruption.

Novelty involves the introduction of new ideas or the recombination of existing fragments of knowledge (Schumpeter, 1939). Scientific and technological advancements do not arise spontaneously but are derived from the existing corpus of knowledge (Arthur, 2009). Novel ideas drive advancements across diverse fields and shape the contours of modern society (Azoulay, Graff Zivin, & Manso, 2011). It also holds profound implications for society, fostering economic growth, technological advancement, and societal progress (Yin, Dong et al., 2022). Scholars have offered various conceptualizations of novelty, emphasizing its multidimensional nature and the interplay of factors influencing its emergence (Azoulay et al., 2011; Uzzi et al., 2013). Notably, promoting diversity and inclusivity within the scientific community can enhance novelty by tapping into a broader pool of talent and perspectives (Lu, Zhang et al., 2022; Xu, Liu et al., 2024). Building upon Schumpeter's seminal work on innovation as the introduction of new products, processes, or forms (Schumpeter, 1939), the identification of atypical knowledge combinations (Uzzi et al., 2013) serves as the widely acknowledged and most suitable means of gauging the novelty of scientific research.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

Disruption manifests through the magnitude of paradigmatic upheaval, delineated by the interconnectedness between past and prospective knowledge (Yang, Gong et al., 2024a; Yang, Hu et al., 2023). Disruptive ideas, grounded in Kuhn's theory of paradigm shifts (Kuhn, 1962), emancipates itself from prevailing paradigms, engendering entirely novel disciplines or paradigms (Wuestman, Hoekman, & Frenken, 2020). The quantification of disruption in science entails an examination of citation links within the knowledge network, elucidating the forward and backward impacts of papers and patents, thereby contributing to an expansive network of knowledge flow. The CD index, introduced by Funk and Owen-Smith (2017), meticulously evaluates the disruptive potential of technologies by scrutinizing the structural attributes of the deep citation network. Wu, Wang, and Evans (2019) and Park et al. (2023) have extended the application of the CD index to scientific literature, furnishing a quantitative metric for gauging the disruptive essence of knowledge.

2.2. Gender Difference in Scientific Impact

A persistent gender disparity exists in scientific authorship, particularly within STEM disciplines and at renowned institutions (Holman, Stuart-Fox, & Hauser, 2018). Career stage and attrition from scientific careers can influence the observed gender gap in research productivity (Huang, Gates et al., 2020). Female researchers are disproportionately likely to encounter authorship disagreements and have their contributions undervalued by colleagues of both genders (Ni, Smith et al., 2021). This underrepresentation of women in science manifests across multiple facets of academic life. Extensive research has documented gender disparities in hiring practices (Moss-Racusin et al., 2012), compensation (Ginther, 2003), access to research funding (Witteman et al., 2019), and recognition through prestigious awards (Ma et al., 2019). Women are demonstrably underrepresented in overall scientific output (Kong et al., 2022) and particularly in senior authorship positions (Ductor et al., 2023; Ross et al., 2022). Furthermore, gender bias has been documented within the peer review process (Witteman et al., 2019).

While prior research suggests research quality itself may not be the primary driver of the career-level gender citation gap (Ferber & Brün, 2011), a growing body of evidence highlights alternative explanations. Studies indicate that disadvantages in career progression faced by women in academia may be a root cause (Ferber & Brün, 2011). Witteman, Haverfield, and Tannenbaum (2021) propose that caregiving responsibilities disproportionately burden female scientists, potentially hindering their overall productivity.

Furthermore, Zhou, Chai, and Freeman (2024) introduce the concept of "gender homophily" in citations, demonstrating a bias towards same-gender authorship. Their findings suggest women-led articles receive fewer citations from subsequent male-led publications, but more citations from subsequent women-led works. This implies a potential subconscious influence of author gender on citation practices. These observations align with broader concerns regarding authorship inequities. Research by Ni et al. (2021) reveals that women are more likely to experience authorship disputes, suggesting a potential gender bias in credit attribution within scientific collaborations. Huang et al. (2020) demonstrate that although career length discrepancies partially explain the gender gap in publication impact, a productivity difference persists. Additionally, awards received by women tend to be associated with lower monetary value, reduced public attention, and less career advancement potential (Ma et al., 2019). These findings highlight the multifaceted nature of the gender citation gap and the need for further investigation into these systemic issues.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

2.3. Gender Difference in Innovation

The landscape of innovation research currently suffers from a dearth of scholarship exploring potential gender disparities in the pursuit of novel and disruptive ideas (Funk & Owen-Smith, 2017; Uzzi et al., 2013). While the field acknowledges the importance of diversity in fostering creativity (Nielsen, Alegria et al., 2017; Wang & Uzzi, 2022), a deeper understanding of how gender influences the generation and reception of groundbreaking concepts remains elusive.

Hofstra, Kulkarni et al. (2020) provide a compelling starting point, suggesting that underrepresented groups, including female scientists in STEM fields, may exhibit a propensity for higher rates of scientific novelty. However, their work also unveils a concerning trend: These novel contributions are often devalued and ultimately discounted. This manifests in lower rates of uptake by other scholars and a diminished likelihood of translating impactful work into successful scientific careers for gender minorities compared to their majority counterparts.

Further evidence emerges from Yang et al. (2022), who investigated the medical field. Their research demonstrates a significant correlation between gender diversity in research teams and the level of novelty in their publications. Teams with a balanced gender composition produced demonstrably more novel work compared to same-gender teams of equivalent size. This pattern held true across various medical subfields, suggesting a generalizable phenomenon. Similarly, Zhang, Wang et al. (2024) explored the impact of gender composition in physics research teams. Their findings indicate that a higher proportion of female scholars and their active participation within mixed-gender teams positively contribute to the generation of disruptive ideas.

2.4. Research Hypotheses

Gender bias and systemic inequities within academic and research institutions are well documented, with evidence indicating that women face numerous barriers to recognition and advancement in scientific fields. These barriers encompass biases in hiring practices (Moss-Racusin et al., 2012), disparities in funding allocation (Witteman et al., 2019), and unequal opportunities for prestigious awards (Ma et al., 2019). Furthermore, research demonstrates that female scientists often encounter challenges in authorship credit and experience higher rates of authorship disputes (Ni et al., 2021). The concept of "gender homophily" in citations (Zhou et al., 2024), suggests that gender can influence citation practices, with male researchers potentially favoring male-led work over female-led work. These systemic issues collectively support the hypothesis that research articles authored by female scientists may receive less scientific impact compared to those authored by male scientists, reflecting broader patterns of gendered disparities in academic recognition and influence.

Thus, we propose our first hypothesis:

H01: Female-led teams receive less scientific impact than those led by male scientists.

The introduction of new ideas or the recombination of existing knowledge fragments drives novelty (Arthur, 2009). Female scientists, often underrepresented in various fields, may approach problems differently and bring unique perspectives that challenge conventional thinking (Hofstra et al., 2020). This diversity of thought is critical for fostering creativity and generating novel ideas. Empirical evidence supports this notion, indicating that diverse teams, including those with a significant female presence, tend to produce more innovative and original research (Yang et al., 2022). Furthermore, promoting diversity and inclusivity within the

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

scientific community can tap into a broader pool of talent and perspectives, enhancing the novelty of research outcomes. This theoretical framework suggests that female-led research teams, by virtue of their diverse perspectives and potential for unconventional thinking, are likely to produce more novel scientific contributions.

Thus, we propose our second hypothesis:

H02: Female-led teams produce more novel ideas than those led by male scientists

Disruptive innovation involves significant paradigmatic upheaval and the potential to redefine existing knowledge structures. Female scientists, who often face systemic barriers and underrepresentation, may develop unique approaches that diverge from established paradigms, leading to disruptive ideas. Research by Zhang et al. (2024) demonstrates that teams with balanced gender compositions produce more disruptive research outcomes. These findings suggest that female-led research, characterized by distinctive insights and a propensity to challenge the status quo, is likely to be more disruptive in nature.

Thus, we propose our third hypothesis:

H03: Female-led teams produce more disruptive ideas than those led by male scientists

3. DATA

3.1. Publication Data

We utilized the Microsoft Academic Graph (MAG) data set (Wang, Shen et al., 2020), a comprehensive scientific publication database that records bibliographic information, authorship, author affiliations, and citation links for articles. Spanning from 1800 to 2021, MAG encompasses over 200 million documents, including journal articles, conference proceedings, preprints, and various research publications. We focused our analysis on journal articles published between 1980 and 2016. This time frame is chosen due to the scarcity and inaccuracy of author information in earlier periods. Additionally, restricting our sample to journal articles published before 2016 ensures a minimum citation window of 5 years (Wang, 2013).

MAG offers a meticulously curated framework for delineating research fields, using advanced machine learning algorithms. At its primary level, MAG defines 19 overarching research domains, which are further divided into 292 subfields. We limited our data set to STEM fields, where biases against women are particularly significant (Yang et al., 2022). These fields include Geography, Environmental Science, Geology, Engineering, Computer Science, Physics, Mathematics, Materials Science, Chemistry, Biology, and Medicine—all categorized as MAG first-level fields.

Moreover, we restricted our analysis to journal articles with at least five references. This criterion serves two main purposes: First, it allows us to use reference information as a proxy for an article's knowledge sources, which is crucial for calculating novelty scores and the CD index; second, it helps filter out nonresearch articles and incomplete records, such as comments, insights, or letters to the editor.

3.2. Author and Gender Data

MAG teams employed sophisticated methodologies, including machine learning and crowdsourcing, to address the name disambiguation problem (Wang et al., 2020). They utilized web search engines to access publicly available information such as personal websites and

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

curricula vitae, thereby enhancing the accuracy of the disambiguation process. Lin, Frey, and Wu (2023) conducted a sample test, demonstrating 100% accuracy and 84% recall, confirming the effectiveness of MAG's disambiguation methods. This data set also provides comprehensive career data for scientists and their affiliation information.

To examine gender differences among team leaders, only the first and last authors are considered, with the last author typically serving as the corresponding author in STEM fields. To ensure data accuracy, scientists with no more than 500 publications are included, minimizing potential errors in name disambiguation (Yang & Wang, 2024). The home field for each scientist is determined based on their first-authored and last-authored publications, assigning the most frequently occurring first-level field as their primary field. Each scientist is assigned only one primary field.

We applied a statistical model (Van Buskirk, Clauset, & Larremore, 2023) to estimate author gender based on first names using data from the MAG database. The model utilizes a cultural consensus model of name-gender associations to infer author gender probabilities. For our data set, probabilities greater than 0.5 indicate female authors, while probabilities below 0.5 indicate male authors. It is important to note that the analysis treats the first author and last author information separately. We included only authors with a minimum of 10 years of career length and at least two first- or last-authored publications in our data set. This criterion excludes individuals who exit the academic field early in their careers and do not achieve team leadership roles.

4. VARIABLES AND MODELS

4.1. Coarsened Exact Matching

To ensure comparability between female and male scientists before our regression analysis, we employed Coarsened Exact Matching (CEM) (Iacus et al., 2012). This advanced matching method enhances the similarity between female and male scientists at the outset. The CEM algorithm involves three main steps: (a) temporarily coarsening each control variable X for matching purposes; (b) sorting all observations into strata with identical values of the coarsened X; and (c) pruning units from the data set that do not include at least one treated and one control unit in any stratum. We performed a one-to-one match, pairing each female scientist with a male scientist, ensuring numerical balance in our sample.

We used a set of predetermined features to control for individual differences, scientific achievements, demographic characteristics, and academic environments. These features included primary fields at level-0, career start year[4], career length, average team sizes, average reference counts, and affiliation rank[5]. We excluded variables such as authors' citation count or productivity from the matching process because these features vary over time and were accounted for in the regression models.

Out of 1.1 million scientists in our sample, we successfully matched 180,184 female scientists with 180,184 male scientists. An examination of the characteristics of the two groups revealed that, prior to publication, they were statistically indistinguishable across all measured dimensions (Figure 2).

[4] We define the start of an author's career as the year they published their first paper in the Microsoft Academic Graph (MAG).

[5] The affiliation rank is determined by the total citation counts of all the papers associated with each affiliation in the Microsoft Academic Graph (MAG).

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

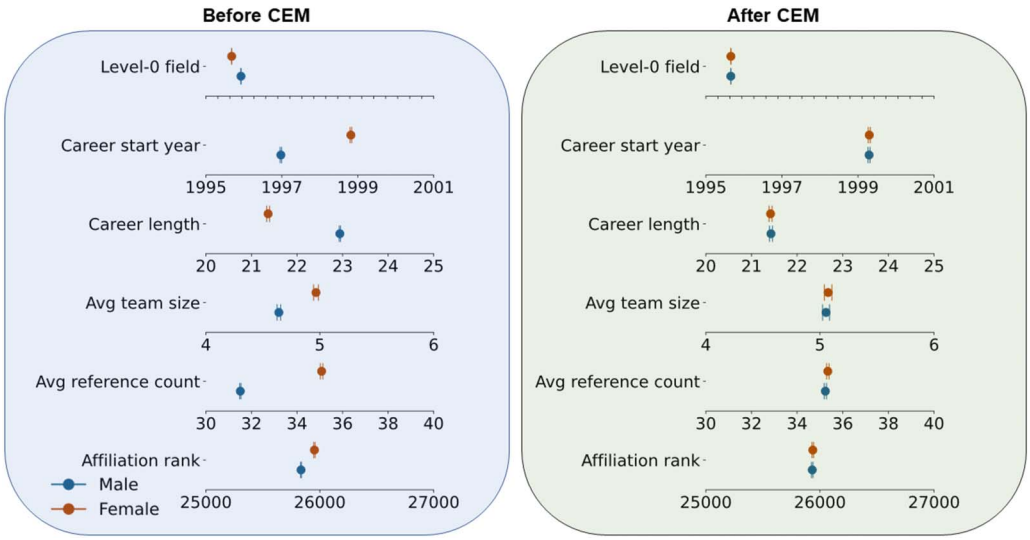


Figure 2. Before-CEM and after-CEM feature comparisons between the female and male group. We compared six different characteristics. The features are defined as follows (from top to bottom): primary fields at MAG level-0, career start year, career length, average team sizes, average reference counts, and affiliation rank (ranked by the total citations in MAG). We see no significant difference between the two groups across any dimension we measured after CEM; Error bar represents the 95% confidence interval.

4.2. Variables

Citation-based metrics of impact are influenced by numerous factors, including the varying citation dynamics across different papers (Aksnes, Piro, & Fossum, 2023), temporal fluctuations in average citation counts, dependencies on specific subfields (Fortunato, Bergstrom et al., 2018), and a highly skewed distribution. To address these variabilities, we use the probability of producing hit papers (or home runs) as a measure of scientific impact (Wang, Jones, & Wang, 2019; Yang, Zhao, & Deng, 2024c). Here, "hit papers" are defined as those in the top 1% of highly cited papers within their respective publication year and field. This approach ensures that the essential information is retained, despite the simplification inherent in using a binary metric.

Novelty involves the introduction of new ideas or the recombination of existing fragments of knowledge (Shi & Evans, 2023; Uzzi et al., 2013). Scientific advancements do not arise spontaneously but are derived from the existing corpus of knowledge (Arthur, 2009). In this study, we followed the methodology of Uzzi et al. (2013), using the yearly journal distribution in the reference list of papers to calculate scientific novelty. As our data set includes papers with at least five references, this method effectively measures atypical combinations of knowledge. We ran Monte Carlo simulations by randomly switching the reference links between random paper pairs while controlling for time. The simulations involve creating reshuffled networks with random edge reassignment while preserving the temporal and distributional attributes of the original citation network. Each journal pairing is transformed into *Z*-scores, representing standardized values. The computation of atypical combinations for each journal pair (pair_{mn}) is encapsulated by

$$Z\ \text{score}_{m,n} = \frac{\text{obs}(\text{pair}_{mn}) - \exp(\text{pair}_{mn})}{\sigma(\text{pair}_{ij})},$$

where $\text{obs}(\text{pair}_{mn})$ signifies the observed frequency of the journal pair in the actual data set, $\exp(\text{pair}_{mn})$ represents the mean, and $\sigma(\text{pair}_{mn})$ denotes the standard deviation of journal pairs obtained from 10 randomized simulations of the reshuffled network. To encapsulate the

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

information within the distribution, denoted as a set $[Z\ \text{score}_{m,n} \mid m, n \in J]$, wherein J encompasses all journals within the reference list, we leverage the 10th percentile value of the novelty set as a succinct summary statistic. We denote articles as novel if their 10th percentile value of the novelty set is lower than 0; otherwise, they are not considered novel articles.

$$Novel\ paper = \begin{cases} 1, & \text{if } 10\text{pct}\left[Z\ \text{score}_{m,n} | m, n \in J\right] < 0 \\ 0, & \text{if } 10\text{pct}\left[Z\ \text{score}_{m,n} | m, n \in J\right] \geq 0 \end{cases}. \quad (1)$$

To evaluate the disruptive nature of academic articles, we employed the CD index (Funk & Owen-Smith, 2017), which assesses structural features within the deep citation network. The CD index is calculated as follows:

$$\text{CD index} = p_D - p_C = \frac{n_i - n_j}{n_i + n_j + n_k} \quad (2)$$

where n_i represents the number of future papers citing the focal paper (FP) that do not reference its cited sources, n_j denotes the number of future papers citing FP and its references, and n_k represents the number of papers citing FP's references without citing FP itself. A lower CD index indicates alignment with established knowledge, while a higher CD index signifies papers with the potential to induce paradigm shifts.

However, the CD index has notable limitations in evaluating the disruptive potential of papers (Leibel & Bornmann, 2023; Yang et al., 2024a). One significant concern is its distribution, which is highly centered around zero (Yang & Deng, 2024; Yang, Yan et al., 2024b). Additionally, the CD index can be largely biased by the unstable parameter n_k. To address these issues, we follow existing studies by employing a dummy variable (Lin et al., 2023), considering a paper to be disruptive if its CD index is above zero. This approach mitigates the bias introduced by the distribution and the unstable parameter n_k.

4.3. Empirical Specification

To address the inherent heterogeneity across scientific disciplines, we incorporated second-level fields of study as fixed effects in our models. We also included year fixed effects (ranging from 1980 to 2016) to account for temporal variations and trends. Additionally, journal fixed effects were introduced to address heterogeneity related to publication outlets, encompassing a total of 39,893 distinct journals. Lastly, we controlled for affiliation fixed effects to account for the leading author's affiliations.

While the combination of CEM and four-level fixed effects addresses most confounding factors, we further included several covariates as additional controls: (a) team size (Wuchty, Jones, & Uzzi, 2007), international collaborations (dummy variable) (Lee, Kogler, & Lee, 2019), and multidisciplinary teams (dummy variable) (Liu, Bu et al., 2024b); (b) grant funding (dummy variable) (Yang, 2024b) and the number of references in the focal papers (Yang, 2024a); and (c) the career age of the author, the author's productivity, the cumulative citations of the author prior to the publication year of the focal paper, and focal field (a dummy variable indicating whether the focal paper is the author's home field). Detailed descriptions of these control variables are provided in Table S1 in the Supplementary material.

We employed a logistic regression model to estimate the effect of having a female scientist as the first or last author on the probability of producing hit papers. Equation 3 is specified as follows:

$$\ln\left(\frac{P(\text{Hit}_i)}{1 - P(\text{Hit}_i)}\right) = \alpha + \beta_1(Female_i) + \sum_{k=2}^{K} \beta_k\left(Control_{k,i}\right) + Y_t + F_f + J_j + A_a + \epsilon_i \quad (3)$$

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

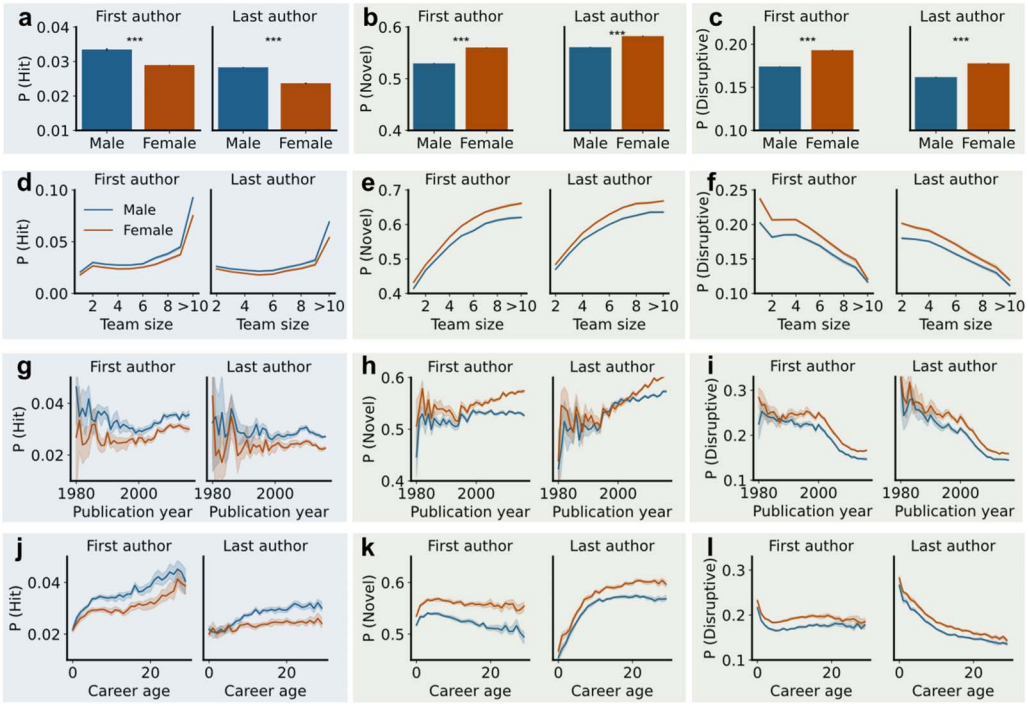


Figure 3. The differences in scientific impact, novelty, and disruption for female-led teams vs. male-led teams. The likelihood of producing (a, d, g, j) high-impact papers, (b, e, h, k) novel papers, and (c, f, i, l) disruptive papers for female-led teams versus male-led teams, considering whether the lead author is the first or the last author. Error bars and shaded areas indicate 95% confidence intervals. *** indicates $p < 0.001$ for a two-tailed Welch's *t*-test.

where i represents the article, t the year of publication, f the field, $Female_i$ is a binary variable indicating if the first author or the last author of paper i is female, $Control_{k,i}$ represents the control variables, Y_t denotes year fixed effects, F_f captures second-level field fixed effects, ε_i denotes the error term and α denotes the constant term. We applied this same model specification to estimate the effects of having a female leader on novelty (the probability of novel papers) and disruption (the probability of disruptive papers).

5. RESULTS

5.1. Female-Led vs. Male-Led Teams

Based on the CEM results, we identified female and male scientists, along with the papers produced by their respective teams. We began by comparing the work produced by female-led and male-led teams, specifically examining their likelihood of generating hit papers, novel papers, and disruptive papers. This straightforward comparison, devoid of any control variables, provides an intuitive understanding of the outcomes.

The results reveal that female-led teams are less likely to produce hit papers[6] compared to their male-led counterparts (diff = 0.45%–0.46%, *p*-values < 0.001 for two-tailed Welch's *t*-test, Figure 3a), irrespective of whether the female scientist is the first or last

[6] Although we defined hit papers as the top 1% of highly cited papers within their respective publication year and field, our data sample shows an average probability of hit papers above 0.01. This is due to our focus on authors with sufficient career length and productivity, resulting in a sample with an average probability of hit papers higher than 0.01.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

Table 1. Effect of female leadership on the probability of hit papers (logit regression)

	(1)	(2)	(3)	(4)	(5)	(6)
	***P* (Hit papers)**					
Models	**First author**			**Last author**		
Female	−0.1465*** (0.0064)	−0.1021*** (0.0072)	−0.0697*** (0.0081)	−0.1814*** (0.0069)	−0.0742*** (0.0077)	−0.0336*** (0.0091)
Paper controls		Yes	Yes		Yes	Yes
Team controls		Yes	Yes		Yes	Yes
Author controls			Yes			Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE		Yes	Yes		Yes	Yes
Affiliation FE			Yes			Yes
Obs.	3,304,856	2,682,999	2,098,780	3,505,475	2,819,138	2,095,369
Pseudo R^2	0.0095	0.2579	0.2768	0.0112	0.2590	0.2777

Note: Robust standard errors are reported in parentheses.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

author. Conversely, teams led by female scientists demonstrate a higher probability of producing novel (diff = 2.2%–3.1%, *p*-values < 0.001, Figure 3b) or disruptive ideas (diff = 1.9%–3.1%, *p*-values < 0.001, Figure 3c), regardless of their authorship position. These findings are robust and consistent across various team sizes (Figures 3d–f), publication years (Figures 3g–i), and stages of career progression (Figures 3j–l).

5.2. Logit Regression Results

We utilized multivariate fixed-effect logistic regression models to investigate gender differences in scientific impact, novelty, and disruption. Initially, we included year and field fixed effects in Model 1–4, followed by the addition of paper and team-level control variables (e.g., grant funding, reference count, team size, multidisciplinary team, international team, journal fixed effects) in Model 2–5. Finally, we incorporated author-level control variables (e.g., career age, author productivity, author citations, focal field, affiliation fixed effects) in Model 3–6, our comprehensive model. The basic results are reported in Tables 1–3, and the detailed results with the coefficients of control variables are shown in Tables S3.1–S3.3 in the Supplementary material.

Our findings indicate that female-led teams are more likely to produce novel and disruptive papers but tend to receive fewer citations. Across all models, the presence of female scientists as first or last authors consistently showed a significantly negative effect on scientific impact (probability of producing high-impact papers) but a significantly positive effect on the probability of generating novel and disruptive papers.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

Table 2. Effect of the female leadership on the probability of novel papers (logit regression)

	(1)	(2)	(3)	(4)	(5)	(6)
	***P* (Novel papers)**					
Models	**First author**			**Last author**		
Female	0.1078*** (0.0023)	0.0302*** (0.0026)	0.0327*** (0.0030)	0.0909*** (0.0022)	0.0400*** (0.0025)	0.0458*** (0.0030)
Paper controls		Yes	Yes		Yes	Yes
Team controls		Yes	Yes		Yes	Yes
Author controls			Yes			Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE		Yes	Yes		Yes	Yes
Affiliation FE			Yes			Yes
Obs.	3,304,856	3,282,703	2,650,210	3,505,475	3,484,394	2,677,601
Pseudo R^2	0.02799	0.14939	0.15343	0.02075	0.14145	0.14896

Note: Robust standard errors are reported in parentheses.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

In our full models, female scientists as first or last authors exhibited a 6.7% (exp(−0.0697) − 1) and 3.3% (exp(−0.0336) − 1) decrease, respectively, in the odds ratio of producing high-impact papers compared to their male counterparts. Conversely, female scientists as first or last authors showed a 3.3% (exp(0.0327) − 1) and 4.7% (exp(0.0458) − 1) increase, respectively, in the odds ratio of producing novel papers. Additionally, female scientists as first or last authors had a 6.7% (exp(0.0649) − 1) and 2.8% (exp(0.0279) − 1) increase, respectively, in the odds ratio of producing disruptive papers compared to male counterparts.

6. HETEROGENEITY AND ROBUST ANALYSIS

6.1. Career Stage and Team Size

Figure 3 suggests that the disparity in the hit paper rate between female-led and male-led teams is not significant at the early career stage but becomes more pronounced in the middle and later stages of their careers. To test this, we conducted split-sample regressions based on the author's career age.

Figure 4 reveals a consistent pattern in gender differences regarding scientific impact, novelty, and disruption across various career stages. Figures 4a and 4b show that female scientists, as first authors, consistently receive fewer citations than their male counterparts at different career stages (except in the first year), with the negative coefficient increasing with career age. When female scientists are the last authors, the gender difference in scientific impact is insignificant in the first 5 years but remains significantly negative in later career stages.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

Table 3. Effect of the female leadership on the probability of disruptive papers (logit regression)

	(1)	(2)	(3)	(4)	(5)	(6)
	P (Disruptive papers)					
Models	First author			Last author		
Female	0.1391*** (0.0029)	0.0632*** (0.0032)	0.0649*** (0.0038)	0.1221*** (0.0029)	0.0360*** (0.0032)	0.0279*** (0.0039)
Paper controls		Yes	Yes		Yes	Yes
Team controls		Yes	Yes		Yes	Yes
Author controls			Yes			Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE		Yes	Yes		Yes	Yes
Affiliation FE			Yes			Yes
Obs.	3,181,201	3,167,066	2,567,637	3,389,438	3,374,407	2,599,804
Pseudo R^2	0.0350	0.1458	0.1558	0.0321	0.1405	0.1494

Note: Robust standard errors are reported in parentheses.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

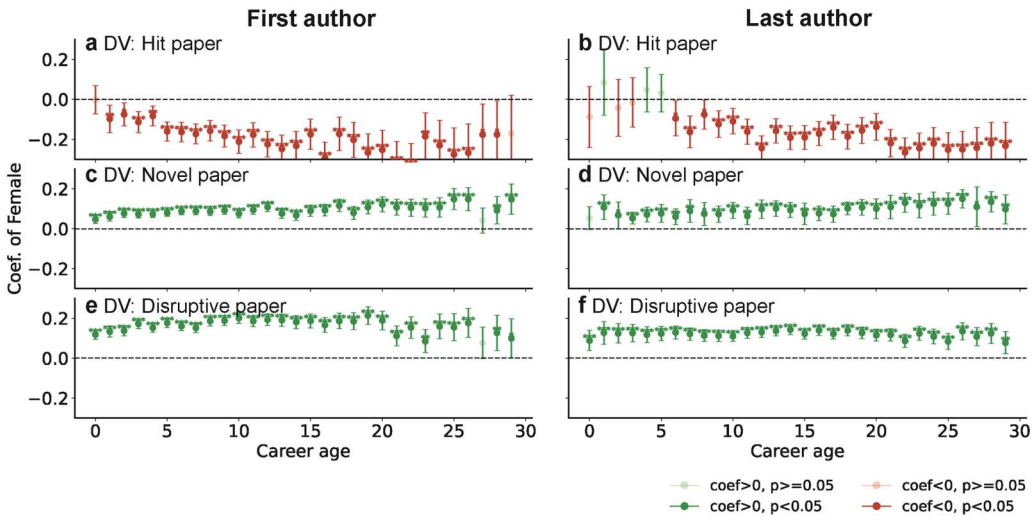


Figure 4. Female-led vs. male-led teams in scientific impact, novelty, and disruption across different career ages. Scatters represent the average effect under different career ages. (a, b) The logistic regression coefficients for female-led teams on the probability of producing hit papers. (c, f) The logistic regression coefficients for female-led teams on the probability of producing novel and disruptive papers. Error bars indicate 95% confidence intervals based on robust standard errors. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

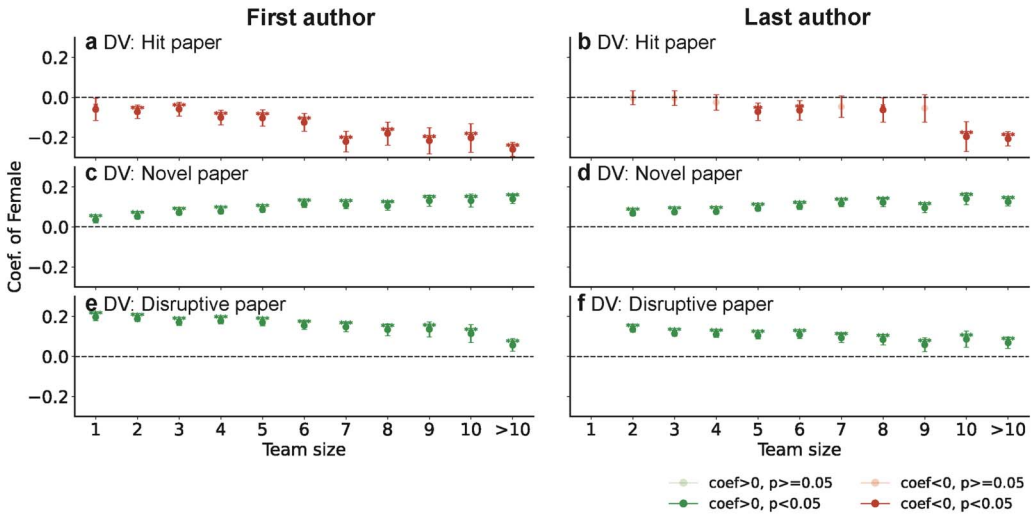


Figure 5. Female-led vs. male-led teams in scientific impact, novelty, and disruption across different team sizes. Scatters represent the average effect under different team sizes. (a, b) The logistic regression coefficients for female-led teams on the probability of producing hit papers. (c–f) The logistic regression coefficients for female-led teams on the probability of producing novel and disruptive papers. Error bars indicate 95% confidence intervals based on robust standard errors. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Figures 4c–f illustrate a consistently significant positive effect of being a female scientist on the probability of producing novel and disruptive papers, for both first and last authors.

These findings indicate that teams led by female scientists consistently produce more novel and disruptive papers throughout their academic careers. However, the bias against their scientific impact becomes more significant in the later stages of their careers.

We further examined whether the differences in outcomes between female-led and male-led teams vary by team size. As shown in Figure 3, the disparity in the rates of high-impact papers between female-led and male-led teams appears larger in larger teams. However, given that larger teams generally have higher hit rates and lower disruptive potential, we conducted split-sample regressions for a more rigorous analysis.

Figure 5 illustrates that the coefficient for female-led teams on high-impact paper rates increases with team size, shifting from insignificant in smaller teams to negatively significant in larger teams. This trend suggests a pronounced bias against female scientists in larger teams, where papers led by female scientists in STEM fields receive fewer scientific impact compared to those led by male scientists.

Regarding novelty and disruption, the coefficients for female leadership on the probability of producing novel and disruptive papers remain consistently and significantly positive across all team sizes. This indicates that female-led teams, regardless of size, are more likely to generate novel and disruptive ideas in STEM fields.

6.2. Heterogeneity Across Fields

We further analyzed the performance of female-led and male-led teams across different fields using split sample regression. Figure 6 illustrates the heterogeneity across MAG level-0 fields, arranged from smallest to largest based on the total number of papers in MAG.

Contrary to our expectations, the bias against female-led teams regarding scientific impact is not significant in many fields. For instance, when a female scientist is the first author in

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

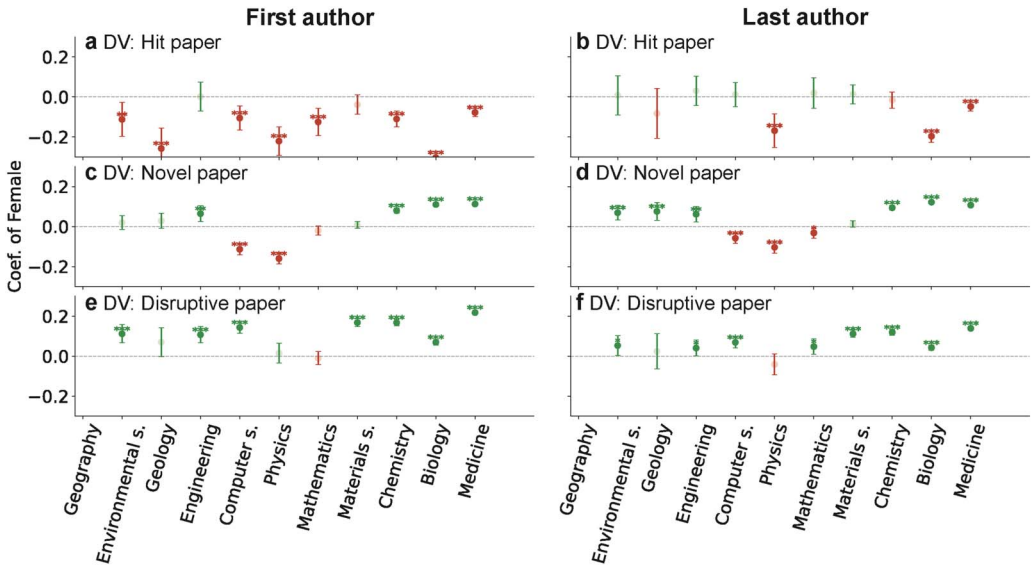


Figure 6. Female-led vs. male-led teams in scientific impact, novelty, and disruption across different fields of study. Scatters represent the average effect under different STEM fields. (a, b) The logistic regression coefficients for female-led teams on the probability of producing hit papers. (c–f) The logistic regression coefficients for female-led teams on the probability of producing novel and disruptive papers. Error bars indicate 95% confidence intervals based on robust standard errors. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Engineering and Material Science, the coefficient is insignificant. Similarly, when a female scientist is the last author, most fields show insignificant results, except for Physics, Biology, and Medicine, where the coefficients are significantly negative. On the other hand, female-led teams do not exhibit significantly higher novelty than male-led teams in Computer Science, Physics, and Mathematics. These findings suggest that the observed bias against female-led teams is predominantly present in Biology and Medicine, the two largest fields of study in MAG.

6.3. Robustness Analysis

We conducted a series of additional analyses to verify the robustness of our findings. First, a significant concern regarding our results is the accuracy of the gender inference procedure, particularly for Asian authors, when using machine-learning-based methods. To address this issue, we conducted a subsample analysis exclusively with U.S. authors, following the methodology of recent studies (Zhou et al., 2024). The results of this analysis, presented in Tables S6.1–S6.3 in the Supplementary material, corroborate our main findings.

We further examined gender differences in scientific impact, novelty, and disruption by analyzing data from scientists in each country separately. Our sample was partitioned based on the countries of the affiliations of the focal authors. The results are depicted in Figure 7, where regions or countries are color-coded: Dark red indicates significantly positive coefficients, light red signifies insignificantly positive coefficients, dark green represents significantly negative coefficients, and light green denotes insignificantly negative coefficients. In most Western countries, the effect of female-led teams on scientific impact is negative. Conversely, the positive effect of female-led teams on the likelihood of producing novel and disruptive papers is significant in most countries. However, the results for Asian countries are less consistent.

Second, we analyzed the impact of mixed leadership in research teams with more than one member. We categorized the sample into four groups based on the gender of the first and last

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

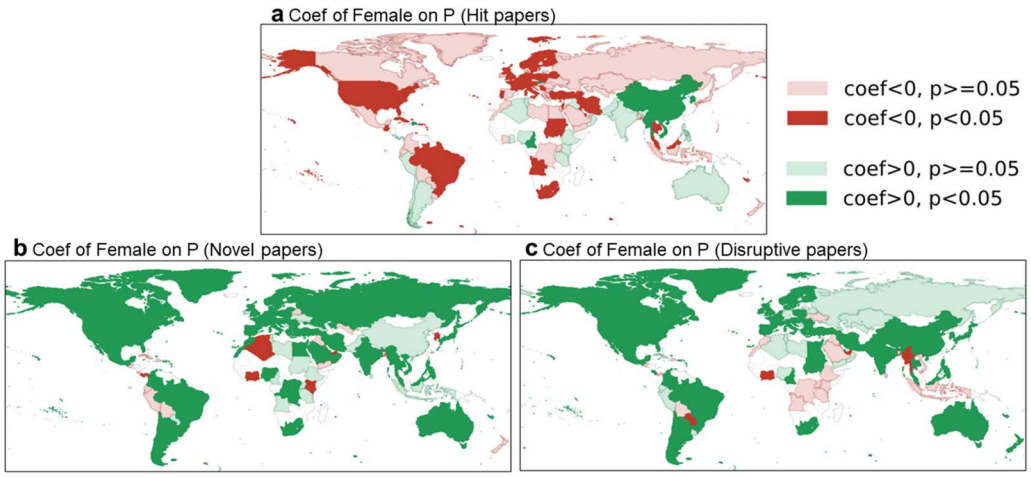


Figure 7. Female-led vs. male-led teams in scientific impact, novelty and disruption across different countries. (a) The Logistic regression coefficient for female leadership on the probability of producing hit papers in each country. (b, c) The logistic regression coefficient for female leadership on the likelihood of producing novel and disruptive papers in each country.

authors: (a) first author male and last author male, (b) first author male and last author female, (c) first author female and last author male, and (d) first author female and last author female. We then compared the average probabilities of producing hit papers, novel papers, and disruptive papers across these groups, as illustrated in Figure 8. Our analysis revealed that teams with both the first and last authors being female consistently had the lowest probability of producing hit papers but the highest probability of producing novel and disruptive papers. Conversely, teams with both the first and last authors being male exhibited the highest probability of producing hit papers but the lowest probability of producing novel and disruptive papers. Teams with mixed-gender leadership (first author male and last author female, and first author female and last author male) fell in between these extremes, displaying similar results. To

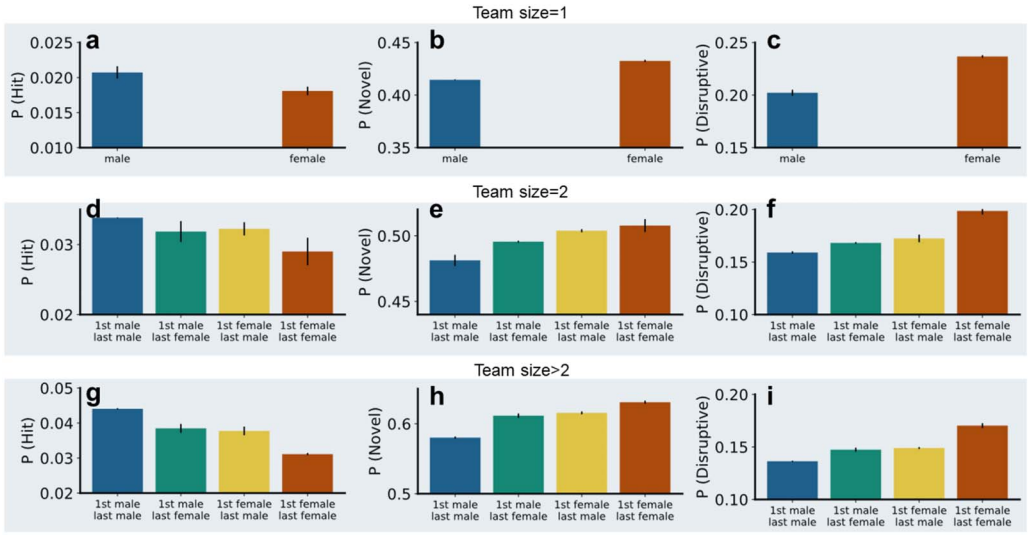


Figure 8. Comparative analysis of mixed-gender leadership in scientific impact, novelty, and disruption across varied team sizes. The likelihood of producing (a, d, g) high-impact papers, (b, e, h) novel papers, and (c, f, i) disruptive papers for each group. For teams with more than one member, the sample was divided into four categories based on the gender of the first and last authors. Error bars represent 95% confidence intervals.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

further substantiate our findings, we conducted a regression analysis using the group with first author male and last author male as the baseline. The regression results, presented in Tables S4.1–4.3 in the Supplementary material, corroborate our observations.

Third, prior studies also indicate that gender-diverse teams often generate more novel and disruptive ideas (Yang et al., 2022). Our analysis, however, does not focus on gender bias, but rather on female leadership as the first or last author, specifically whether the research is led by a female. To isolate the impact of gender diversity on the results, we construct a variable called gender entropy, defined as gender entropy = $-p_f \ln(p_f) - (1-p_f) \ln(1-p_f)$, where p_f represents the probability of female authors in team. The value of gender entropy ranges from 0 to 1. Higher gender entropy indicates higher diversity. We incorporate gender entropy into the regression models. The results, presented in Tables S6.1–S6.3 in the Supplementary material, demonstrate that our findings largely persist even when controlling for gender diversity, with the exception of female last authorship on scientific impact. This suggests that gender diversity does not fully account for our findings: Female leadership is associated with lower scientific impact, higher novelty, and higher disruption.

In Tables S7.1 and S8.1 in the Supplementary material, we validated our analysis by employing alternative measures of scientific impact, including the top 10% most highly cited papers and the 5-year citation count. In Table S9.1 in the Supplementary material, we utilized alternative measures of novelty. Additionally, we used the raw CD index and the five-year CD index as dependent variables in Tables S10.1 and S11.1 in the Supplementary material. The results consistently align with our primary findings, collectively underscoring the robustness and reliability of our conclusions.

7. DISCUSSIONS

7.1. Implications

Our study contributes to the existing literature by shedding light on significant gender disparities in scientific impact, novelty, and disruption within STEM fields. The finding that female-led teams produce more novel and disruptive ideas yet generate less scientific impact than their male counterparts. This underscores a systemic bias that undervalues the impact of female scientists' work.

Theoretical frameworks such as the "Matthew Matilda effect" (Rossiter, 1993) and "female penalty for novelty" (Trapido, 2022) provide a lens through which to interpret these findings. Established male researchers, who dominate the scientific publication landscape, continue to receive a disproportionate share of citations and recognition, despite evidence suggesting comparable or superior contributions by female scientists in terms of novelty and disruption. This phenomenon suggests that current recognition metrics may not adequately capture the quality of scientific contributions from diverse groups, thereby perpetuating gender disparities in scientific recognition and advancement. According to the perspective of female penalty for novelty, male scientists are generally expected to produce more innovative research compared to their female counterparts. When women do create novel work, it is often viewed as a departure from gender norms. Consequently, due to these gender stereotypes, the unique research output of female scientists may not receive equal evaluation or recognition as that of their male peers. Research findings indicate that women tend to receive less acknowledgment than men for their innovative contributions across different sectors, including academia (Kabat-Farr & Cortina, 2012; Schmutz & Faupel, 2010; Trapido, 2022). In a prior study conducted by the authors, it was observed that female PhD graduates tend to

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

exhibit a lower level of novelty in their doctoral theses compared to their male counterparts (Liu, Xie et al., 2024a). The disparities in these findings could be linked to variations in scientific publication types, career stages, and disciplines. Further exploration should be performed to understand the contrasting directions of gender discrepancies in novelty between doctoral theses and journal articles.

Furthermore, the study highlights the role of team size in exacerbating gender biases. We find that the negative effect of female leadership on scientific impact is more pronounced in larger teams. This suggests that biases against female scientists are amplified in environments where research output is more visible and competitive, potentially affecting career advancement and funding opportunities.

Practically, our study calls for policy interventions aimed at mitigating gender biases in the evaluation and recognition of scientific contributions. Institutional policies and practices should be revisited to ensure equitable evaluation metrics that consider the diversity of scientific contributions and the contexts in which they are made. This includes reevaluating citation practices to minimize bias against female-authored research and ensuring that contributions from diverse groups are appropriately recognized and valued. Additionally, our findings suggest that efforts to promote diversity and inclusion within scientific teams are not only ethically imperative but also beneficial for scientific innovation. Policies that support the recruitment, retention, and advancement of women in STEM fields, particularly in leadership and decision-making roles, are crucial. Initiatives that provide mentorship, networking opportunities, and support for work-life balance can help mitigate the barriers faced by female scientists and contribute to a more inclusive and equitable scientific community.

7.2. Limitations and Future Avenues

Despite our comprehensive analysis, this study is not without limitations. First, our study focused on the first and last author positions as proxies for scientific leadership, potentially overlooking the contributions of coauthors and collaborators. Future research could explore the role of middle authors and their impact on gender disparities in scientific recognition. Second, our analysis primarily examines gender differences in scientific impact, novelty, and disruption, without delving into the underlying mechanisms of these disparities. Further qualitative research could help uncover the specific barriers and biases that contribute to the underestimation of female scientists' contributions. Third, while we account for various confounding factors in our CEM procedures and regression models, there may be unobserved variables that influence our findings. Future studies could employ alternative methodological approaches, such as machine learning algorithms or causal inference methods, to address these potential confounders more effectively. Finally, while our study calls for policies to counteract biases and promote equity in scientific evaluation, the effectiveness of such policies remains an area for future investigation. Policy evaluations could provide insights into the impact of diversity initiatives on reducing gender disparities in scientific recognition.

ACKNOWLEDGMENTS

The authors deeply appreciate the constructive comments from the reviewers.

AUTHOR CONTRIBUTIONS

Alex J. Yang: Conceptualization; Data curation; Methodology; Resources; Software; Visualization; Writing – original draft. Ying Ding: Supervision; Validation; Writing – review & editing. Meijun Liu: Funding acquisition; Supervision; Writing – original draft; Writing – review & editing.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

COMPETING INTERESTS

The authors have no competing interests.

SUPPORTING INFORMATION

The Supplementary Material is available at https://github.com/AlexJieYang/Gender_Innovation.

FUNDING INFORMATION

This paper is supported by the Open Fund for Innovative Evaluation from Fudan University (#CXPJ2024004), the Youth Program of National Natural Science Foundation in China (No: 72104007), Shanghai Pujiang Program (No: 21PJC026) and Key Project of the National Natural Science Foundation of China (No: 72234001).

DATA AVAILABILITY

The Microsoft Academic Graph data can be downloaded via https://zenodo.org/record/2628216#.Y-7RR_5Bz-g. Other data used in this study can be obtained by making reasonable requests. The code is available at https://github.com/AlexJieYang/Gender_Innovation. Note that one may need Python-3.10 as well as R-4.3.2 to replicate the code. We also used the code provided by Gates and Barabási (2023) to calculate variables.

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

REFERENCES

Aksnes, D. W., Piro, F. N., & Fossum, L. W. (2023). Citation metrics covary with researchers' assessments of the quality of their works. *Quantitative Science Studies*, *4*(1), 105–126. https://doi.org/10.1162/qss_a_00241

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. New York: Simon & Schuster.

Azoulay, P., Graff Zivin, J. S., & Manso, G. (2011). Incentives and creativity: Evidence from the academic life sciences. *RAND Journal of Economics*, *42*(3), 527–554. https://doi.org/10.1111/j.1756-2171.2011.00140.x

Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, *26*(3), 312–327. https://doi.org/10.1017/pan.2018.12

Ductor, L., Goyal, S., & Prummer, A. (2023). Gender and collaboration. *Review of Economics and Statistics*, *105*(6), 1366–1378. https://doi.org/10.1162/rest_a_01113

Ferber, M. A., & Brün, M. (2011). The gender gap in citations: Does it persist? *Feminist Economics*, *17*(1), 151–158. https://doi.org/10.1080/13545701.2010.541857

Fortunato, S., Bergstrom, C. T., Boerner, K., Evans, J. A., Helbing, D., ... Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185. https://doi.org/10.1126/science.aao0185, PubMed: 29496846

Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science*, *63*(3), 791–817. https://doi.org/10.1287/mnsc.2015.2366

Gates, A. J., & Barabási, A.-L. (2023). Reproducible science of science at scale: *pySciSci*. *Quantitative Science Studies*, *4*(3), 700–710. https://doi.org/10.1162/qss_a_00260

Ginther, D. K. (2003). Is MIT an exception? Gender pay differences in academic science. *Bulletin of Science, Technology & Society*, *23*(1), 21–26. https://doi.org/10.1177/0270467602239767

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*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117, PubMed: 32291335

Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLOS Biology*, *16*(4), e2004956. https://doi.org/10.1371/journal.pbio.2004956, PubMed: 29672508

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*(9), 4609–4616. https://doi.org/10.1073/pnas.1914221117, PubMed: 32071248

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened Exact Matching. *Political Analysis*, *20*(1), 1–24. https://doi.org/10.1093/pan/mpr013

Kabat-Farr, D., & Cortina, L. M. (2012). Selective incivility: Gender, race, and the discriminatory workplace. In S. Fox & T. R. Lituchy (Eds.), *Gender and the dysfunctional workplace* (pp. 120–134). Edward Elgar Publishing. https://doi.org/10.4337/9780857932600.00014

Kong, H., Martin-Gutierrez, S., & Karimi, F. (2022). Influence of the first-mover advantage on the gender disparities in physics citations. *Communications Physics*, *5*, 243. https://doi.org/10.1038/s42005-022-00997-x

Kuhn, T. S. (1962). Historical structure of scientific discovery. *Science*, *136*(3518), 760–764. https://doi.org/10.1126/science.136.3518.760, PubMed: 14460344

Lee, C., Kogler, D. F., & Lee, D. (2019). Capturing information on technology convergence, international collaboration, and knowledge flow from patent documents: A case of information

and communication technology. *Information Processing & Management*, *56*(4), 1576–1591. https://doi.org/10.1016/j.ipm.2018.09.007

Leibel, C., & Bornmann, L. (2023). What do we know about the disruption index in scientometrics? An overview of the literature. *Scientometrics*, *129*, 601–639. https://doi.org/10.1007/s11192-023-04873-5

Lerman, K., Yu, Y., Morstatter, F., & Pujara, J. (2022). Gendered citation patterns among the scientific elite. *Proceedings of the National Academy of Sciences*, *119*(40), e2206070119. https://doi.org/10.1073/pnas.2206070119, PubMed: 36161888

Lin, Y., Frey, C. B., & Wu, L. (2023). Remote collaboration fuses fewer breakthrough ideas. *Nature*, *623*(7989), 987–991. https://doi.org/10.1038/s41586-023-06767-1, PubMed: 38030778

Liu, M., Xie, Z., Yang, A. J., Yu, C., Xu, J., Ding, Y., & Bu, Y. (2024a). The prominent and heterogeneous gender disparities in scientific novelty: Evidence from biomedical doctoral theses. *Information Processing & Management*, *61*(4), 103743. https://doi.org/10.1016/j.ipm.2024.103743

Liu, M., Yang, S., Bu, Y., & Zhang, N. (2023). Female early-career scientists have conducted less interdisciplinary research in the past six decades: Evidence from doctoral theses. *Humanities and Social Sciences Communications*, *10*, 918. https://doi.org/10.1057/s41599-023-02392-5

Liu, M., Zhang, N., Hu, X., Jaiswal, A., Xu, J., ... Bu, Y. (2022). Further divided gender gaps in research productivity and collaboration during the COVID-19 pandemic: Evidence from coronavirus-related literature. *Journal of Informetrics*, *16*(2), 101295. https://doi.org/10.1016/j.joi.2022.101295, PubMed: 35529705

Liu, X., Bu, Y., Li, M., & Li, J. (2024b). Monodisciplinary collaboration disrupts science more than multidisciplinary collaboration. *Journal of the Association for Information Science and Technology*, *75*(1), 59–78. https://doi.org/10.1002/asi.24840

Lu, C., Zhang, C., Xiao, C., & Ding, Y. (2022). Contributorship in scientific collaborations: The perspective of contribution-based byline orders. *Information Processing & Management*, *59*(3), 102944. https://doi.org/10.1016/j.ipm.2022.102944

Ma, Y., Oliveira, D. F. M., Woodruff, T. K., & Uzzi, B. (2019). Women who win prizes get less money and prestige. *Nature*, *565*(7739), 287–288. https://doi.org/10.1038/d41586-019-00091-3, PubMed: 30651627

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109, PubMed: 22988126

Ni, C., Smith, E., Yuan, H., Larivière, V., & Sugimoto, C. R. (2021). The gendered nature of authorship. *Science Advances*, *7*(36), eabe4639. https://doi.org/10.1126/sciadv.abe4639, PubMed: 34516891

Nielsen, M. W., Alegria, S., Börjeson, L., Etzkowitz, H., Falk-Krzesinski, H. J., ... Schiebinger, L. (2017). Gender diversity leads to better science. *Proceedings of the National Academy of Sciences*, *114*(8), 1740–1742. https://doi.org/10.1073/pnas.1700616114, PubMed: 28228604

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, *613*(7942), 138–144. https://doi.org/10.1038/s41586-022-05543-x, PubMed: 36600070

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 men. *Nature*, *608*(7921), 135–145. https://doi.org/10.1038/s41586-022-04966-w, PubMed: 35732238

Rossiter, M. W. (1993). The Matthew Matilda effect in science. *Social Studies of Science*, *23*(2), 325–341. https://doi.org/10.1177/030631293023002004

Schmutz, V., & Faupel, A. (2010). Gender and cultural consecration in popular music. *Social Forces*, *89*(2), 685–707. https://doi.org/10.1353/sof.2010.0098

Schumpeter, J. A. (1939). *Business cycles: A theoretical, historical, and statistical analysis of the capitalist process* (*Vol. 1*). New York: McGraw-Hill.

Shi, D., Liu, W., & Wang, Y. (2023). Has China's Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists? *Science*, *379*(6627), 62–65. https://doi.org/10.1126/science.abq1218, PubMed: 36603081

Shi, F., & Evans, J. (2023). Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications*, *14*, 1641. https://doi.org/10.1038/s41467-023-36741-4, PubMed: 36964138

Teich, E. G., Kim, J. Z., Lynn, C. W., Simon, S. C., Klishin, A. A., ... Bassett, D. S. (2022). Citation inequity and gendered citation practices in contemporary physics. *Nature Physics*, *18*(10), 1161–1170. https://doi.org/10.1038/s41567-022-01770-1

Trapido, D. (2022). The female penalty for novelty and the offsetting effect of alternate status characteristics. *Social Forces*, *100*(4), 1592–1618. https://doi.org/10.1093/sf/soab053

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472. https://doi.org/10.1126/science.1240474, PubMed: 24159044

Van Buskirk, I., Clauset, A., & Larremore, D. B. (2023). An open-source cultural consensus approach to name-based gender classification. *Proceedings of the International AAAI Conference on Web and Social Media*, *17*, 866–877. https://doi.org/10.1609/icwsm.v17i1.22195

Wang, D., & Uzzi, B. (2022). Weak ties, failed tries, and success. *Science*, *377*(6612), 1256–1258. https://doi.org/10.1126/science.add0692, PubMed: 36108030

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872. https://doi.org/10.1007/s11192-012-0775-9

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, *1*(1), 396–413. https://doi.org/10.1162/qss_a_00021

Wang, Y., Jones, B. F., & Wang, D. (2019). Early-career setback and future career impact. *Nature Communications*, *10*, 4331. https://doi.org/10.1038/s41467-019-12189-3, PubMed: 31575871

Witteman, H. O., Haverfield, J., & Tannenbaum, C. (2021). COVID-19 gender policy changes support female scientists and improve research quality. *Proceedings of the National Academy of Sciences*, *118*(6), e2023476118. https://doi.org/10.1073/pnas.2023476118, PubMed: 33531366

Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2019). Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *Lancet*, *393*(10171), 531–540. https://doi.org/10.1016/S0140-6736(18)32611-4, PubMed: 30739688

Wu, L. F., Wang, D. S., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378–382. https://doi.org/10.1038/s41586-019-0941-9, PubMed: 30760923

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099, PubMed: 17431139

Wuestman, M., Hoekman, J., & Frenken, K. (2020). A typology of scientific breakthroughs. *Quantitative Science Studies*, *1*(3), 1203–1222. https://doi.org/10.1162/qss_a_00079

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026

Xu, H., Liu, M., Bu, Y., Sun, S., Zhang, Y., ... Ding, Y. (2024). The impact of heterogeneous shared leadership in scientific teams. *Information Processing & Management*, *61*(1), 103542. https://doi.org/10.1016/j.ipm.2023.103542

Yang, A. J. (2024a). On the temporal diversity of knowledge in science. *Journal of Informetrics*, *18*(4), 101594. https://doi.org/10.1016/j.joi.2024.101594

Yang, A. J. (2024b). Unveiling the impact and dual innovation of funded research. *Journal of Informetrics*, *18*(1), 101480. https://doi.org/10.1016/j.joi.2023.101480

Yang, A. J., & Deng, S. (2024). Dynamic patterns of the disruptive and consolidating knowledge flows in Nobel-winning scientific breakthroughs. *Quantitative Science Studies*, 1–17. https://doi.org/10.1162/qss_a_00323

Yang, A. J., Gong, H., Wang, Y., Zhang, C., & Deng, S. (2024a). Rescaling the disruption index reveals the universality of disruption distributions in science. *Scientometrics*, *129*(1), 561–580. https://doi.org/10.1007/s11192-023-04889-x

Yang, A. J., Hu, H., Zhao, Y., Wang, H., & Deng, S. (2023). From consolidation to disruption: A novel way to measure the impact of scientists and identify laureates. *Information Processing & Management*, *60*(5), 103420. https://doi.org/10.1016/j.ipm.2023.103420

Yang, A. J., Yan, X., Hu, H., Hu, H., Kong, J., & Deng, S. (2024b). Are disruptive papers more likely to impact technology and society? *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.24947

Yang, A. J., Zhao, S. X., & Deng, S. (2024c). Revisiting delayed recognition in science: A large-scale and comprehensive study. *Journal of Information Science*. https://doi.org/10.1177/01655515241244462

Yang, W., & Wang, Y. (2024). Higher-order structures of local collaboration networks are associated with individual scientific productivity. *EPJ Data Science*, *13*, 15. https://doi.org/10.1140/epjds/s13688-024-00453-6

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*(36), e2200841119. https://doi.org/10.1073/pnas.2200841119, PubMed: 36037387

Yin, Y., Dong, Y., Wang, K., Wang, D., & Jones, B. F. (2022). Public use and public funding of science. *Nature Human Behaviour*, *6*(10), 1344–1350. https://doi.org/10.1038/s41562-022-01397-5, PubMed: 35798885

Zhang, M.-Z., Wang, T.-R., Lyu, P.-H., Chen, Q.-M., Li, Z.-X., & Ngai, E. W. T. (2024). Impact of gender composition of academic teams on disruptive output. *Journal of Informetrics*, *18*(2), 101520. https://doi.org/10.1016/j.joi.2024.101520

Zhao, X., Akbaritabar, A., Kashyap, R., & Zagheni, E. (2023). A gender perspective on the global migration of scholars. *Proceedings of the National Academy of Sciences*, *120*(10), e2214664120. https://doi.org/10.1073/pnas.2214664120, PubMed: 36848569

Zhou, S., Chai, S., & Freeman, R. B. (2024). Gender homophily: In-group citation preferences and the gender disadvantage. *Research Policy*, *53*(1), 104895. https://doi.org/10.1016/j.respol.2023.104895

Downloaded from direct.mit.edu/qss/article-pdf/5/4/861/2482658/qss_a_00335.pdf by guest on 06 November 2026