PLOS ONE

The Role of Gender in Scholarly Authorship

Jevin D. West[1]*, Jennifer Jacquet[2], Molly M. King[3], Shelley J. Correll[3], Carl T. Bergstrom[1,4]

1 Department of Biology, University of Washington, Seattle, Washington, United States of America, 2 Environmental Studies, New York University, New York, New York, United States of America, 3 Department of Sociology, Stanford University, Stanford, California, United States of America, 4 Santa Fe Institute, Santa Fe, New Mexico, United States of America

Abstract

Gender disparities appear to be decreasing in academia according to a number of metrics, such as grant funding, hiring, acceptance at scholarly journals, and productivity, and it might be tempting to think that gender inequity will soon be a problem of the past. However, a large-scale analysis based on over eight million papers across the natural sciences, social sciences, and humanities reveals a number of understated and persistent ways in which gender inequities remain. For instance, even where raw publication counts seem to be equal between genders, close inspection reveals that, in certain fields, men predominate in the prestigious first and last author positions. Moreover, women are significantly underrepresented as authors of single-authored papers. Academics should be aware of the subtle ways that gender disparities can occur in scholarly authorship.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jevinw@u.washington.edu

Introduction

Gender inequities and gender biases persist in higher education. After decades of high female enrollment in most PhD fields, women represent one-quarter of full professors and earn on average 80% of the salary of men in comparable positions [1]. A recent report [2] surveyed 1800 faculty across six science and engineering disciplines and found men publish significantly more in chemistry and mathematics, while women publish more in electrical engineering (there were no significant differences found in biology, civil engineering, and physics). A recent experiment tested the role of gender in hiring by asking 127 science faculty to evaluate potential lab manager applications and found faculty gave identical applications higher scores if the applicant had a male name [3]. Another recent analysis of commissioned articles in two prestigious journals published in 2010 and 2011 showed that women scientists are underrepresented; for instance, women wrote just 3.8% of earth and environmental sciences articles for *Nature* News & Views, although they represent 20% of the scientists in this discipline [4]. With the use of alphabetical authorship listings declining over time [5], and given the complexity of evaluating intellectual contributions [6] in increasingly collaborative efforts, understanding patterns of authorship order becomes increasingly important.

Here we use the JSTOR corpus—a body of academic papers from a range of scholarly disciplines spanning five centuries—to examine trends in the gender composition of academic authorship through time. We pay particular attention to authorship order, given that first and sometimes last author publications are at least as important as raw publication counts for hiring, promotion, and tenure, particularly in scientific fields [7]. Studies of authorship in the medical literature reveal, for instance, that women have been historically underrepresented in the prestige positions of first and last author, and that while discrepancies have recently declined in the first author position, women remain underrepresented as last authors [8,9,10,11]. To view authorship patterns in their disciplinarily context, we use a network-based community detection approach to categorize hierarchically each paper in our study corpus. This yields a hierarchical classification of all papers in our study and allows us to study and compare patterns of gender representation in individual fields of any size and scale.

Methods

The JSTOR corpus

The JSTOR corpus (http://www.jstor.org) is a digital archive of published scholarly research that spans the sciences and humanities from 1545 to the present day. At the time of this analysis, the JSTOR corpus comprised 8.3 million documents ranging from 1545 until early 2011, including 4.2 million research articles. Approximately 1.8 million of these documents (97% of which are research articles) cite or are cited by other documents in the JSTOR corpus and thus are amenable to network analysis. We call this group the "JSTOR network dataset". Moreover 94% of these 1.8 million articles are part of a single giant component of the citation network, such that any of these articles can be reached from any other by following citation trails forwards and backwards. We restrict our analysis to the JSTOR network dataset because this is the portion of the JSTOR corpus that we can hierarchically categorize using citation information. For a list of the main fields available in JSTOR dataset, see Table 1. The gender composition of the identified authors in the network dataset (21.9% female) is close to that of the identified authors in the entire corpus (20.8% percent).

Table 1. Gender composition from 1990–2011 for disciplines (i.e., groups at the first level of hierarchical clustering) with at least 5,000 authorships.

Field	% female	authorships
Mathematics	10.64	6134
Philosophy	12.04	12190
Economics	13.68	69142
Probability and Statistics	18.11	28324
Political science - international	19.07	14908
Political science-US domestic	19.09	15705
Ecology and evolution	22.76	279012
Law	24.21	18503
Organizational and marketing	25.44	32119
Physical anthropology	27.05	16296
Radiation damage	27.69	7825
Classical studies	28.88	6372
Molecular & Cell biology	29.25	277032
History	30.47	15585
Veterinary medicine	31.81	10960
Cognitive science	32.12	12786
Anthropology	36.46	19900
Pollution and occupational health	37.57	32108
Sociology	41.41	44895
Demography	41.90	7600
Education	46.35	28635

doi:10.1371/journal.pone.0066212.t001

Mapping the hierarchical structure of scholarly research

The scientific literature can be viewed as a large network in which papers are linked by citation relationships [12]. The topology of scientific networks can be used to map the structure of science, and the map equation [13,14] has proven to be a particularly effective method [15]. However, such maps of science have typically shown only a single layer of structure. To map the

structure of scholarly disciplines, fields and subfields, we turn to the hierarchical map equation [16], which reveals multiple levels of substructure within a network. Using the hierarchical map equation on the network of citations, we create a multi-scale map of the JSTOR network dataset in the form of a hierarchical classification that assigns each paper to a major domain, field, subfield, speciality within subfield, and so forth. For example, Bill Hamillton's classic 1980 paper "Sex versus asex versus parasite" is classified as residing in Ecology and evolution : Population genetics : Sexual and asexual reproduction : Sex and virulence. We used the May 13th, 2012 version of the hierarchical map equation code; improvements to that search algorithm made subsequent to our analysis may find somewhat flatter hierarchies than that reported here. While the algorithm made the decisions about how many fields exist and which papers are assigned to which fields, we manually assigned descriptive names to each field or subfield to facilitate navigation. The names are intended as a general indication of subject matter rather than as a definitive classification.

Determining gender of authors

We use US Social Security Administration records to determine gender from first names. The US Social Security Administration website (http://www.ssa.gov/oact/babynames/) makes available the top 1000 names annually for each of the 153 million boys and 143 million girls born from 1880–2010. (These data acknowledge only two genders.) We assume we can identify an author's gender if the author's first name is associated with a single gender in social security records at least 95% of the time, as with 'Mary', or 'John'. Otherwise, as with 'Leslie' or 'Sidney', we are unable to identify the gender and do not include that author in our analysis. Since in any given era, androgynous names are more likely to be females, this may slightly downwardly bias our estimates of women [17]. Similarly, we are unable to classify names that never appear in the top 1000 for either gender in the US records. As a result, authors of some nationalities may be underrepresented in our data set. In a few rare cases national differences may cause misleading assignments for non-US authors (e.g. 'Andrea' is typically a female name in the US but a male name in Italy). By this method we are able to assign genders to 6879 unique first names: 3809 female and 3070 male.
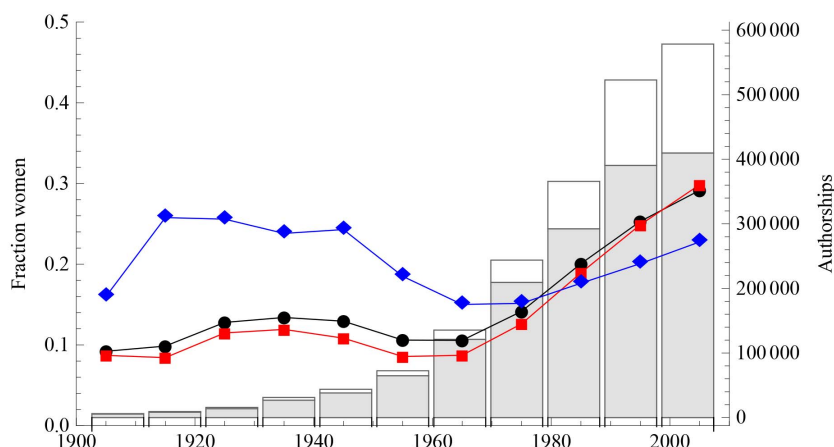
Figure 1. Authorships and gender composition in the JSTOR network dataset, by decade. Shaded bars represent male authorships, unshaded bars represent female authorships. The black line indicates the fraction of authorships that are women, the red line indicates the fraction of first authorships that are women, and the blue line indicates the fraction of last authorships that are women.
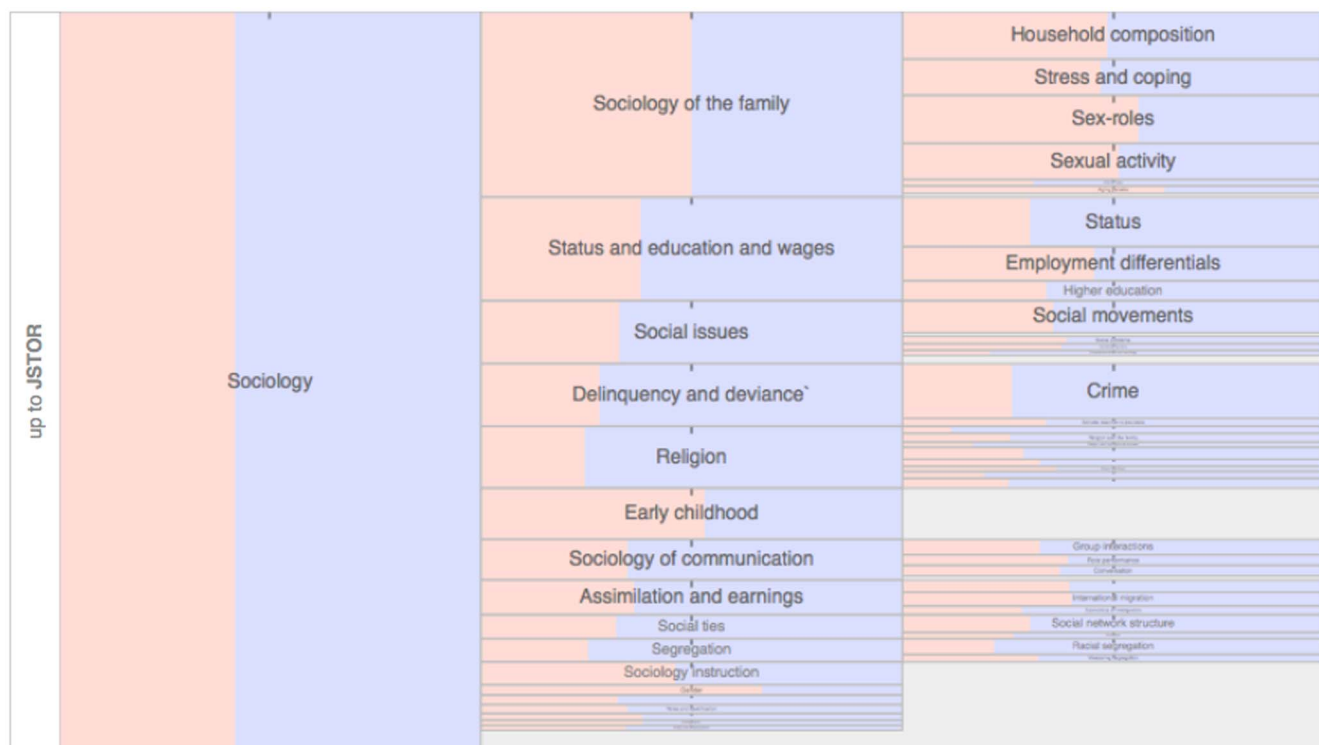doi:10.1371/journal.pone.0066212.g001

Figure 2. Even in fields with a gender composition near parity, men (blue bars) and women (pink bars) are unequally distributed in subfields. Shown here is sociology and its subfields from 1990 to the present. An interactive version of this graph, covering all fields and subfields of the JSTOR network dataset, is available online at http://www.eigenfactor.org/gender/.
doi:10.1371/journal.pone.0066212.g002

We extracted the first names of all authors in the JSTOR network dataset, discarding those authors who list only initials. An *instance of authorship* consists of a person and a paper for which the person is designated as a co-author. There are 3.6 million authorships in the JSTOR network dataset; of these we are able to extract a full first name for 2.8 million authorships (77%) associated with 1.5 million papers. (The exclusion of authors with only first initials may exclude women authors disproportionately, particularly in early eras when women may have been more likely than men to publish with initials to avoid potential discrimination.) Of these 2.8 million authorships with full first names, we are able to confidently assign gender to 73.3%. The remaining authorships involve names not in the US social security top 1000 lists (24.3%), or names associated with both genders (2.4%). The final data analyzed include all papers where we know the gender of one or more authors.

Gender and authorship order

We look at the gender composition of all papers with any number of authors in the JSTOR network dataset. For every field, subfield, and so-forth, we calculate both the overall gender composition and the gender composition of each authorship position—first, second, third, etc. In some fields, such as molecular biology, the last author position of a paper conveys a special meaning: the last author is typically the principal investigator or group leader of multi-author effort. This is especially the case for papers with at least three authors. Therefore we also report the gender frequency in the last-author position for all papers with three or more co-authors. We then compare the gender frequencies at each author position with the overall gender frequency in the same field. If authorship order were gender-

Table 2. Percentage of women relative to total PhDs and percentage of women in tenure or tenure track positions and full professorships in Science and Engineering from 1960–2006 (data from reference [33]) as well as percentage of women in various author positions from 1960–2009 as a result of this analysis.

	1960s	1970s	1980s	1990s	2000s
% PhDs overall	7–9	9–22	23–30	29–37	38–40
Computer sciences	n/a	n/a	9–18	14–19	16–21
Engineering	0–1	1–3	4–8	9–15	16–20
Life Sciences	8–14	13–25	26–39	38–46	47–52
Mathematics	5–7	6–16	13–18	18–24	25–30
Physical Sciences	3–5	5–11	11–19	19–24	25–29
Psychology	18–24	24–41	42–56	58–67	67–71
Social Sciences	8–12	11–26	27–35	33–42	43–46
% Tenure track faculty	n/a	n/a	10–15	16–22	24–28
Full Professors	n/a	5	5–8	9–14	16–19
% Authors overall	10.6	14.2	20.1	25.3	29.2
Single author	8.7	12.5	18.7	24.5	28.5
1st author	9.2	12.9	19.3	25.3	30.9
2nd author	14.8	16.2	20.8	25.0	28.8
Last author	15.0	15.2	17.6	20.1	22.8

1st and 2nd author positions are listed for papers with at least two authors. Last author percentage is for papers with at least three authors.
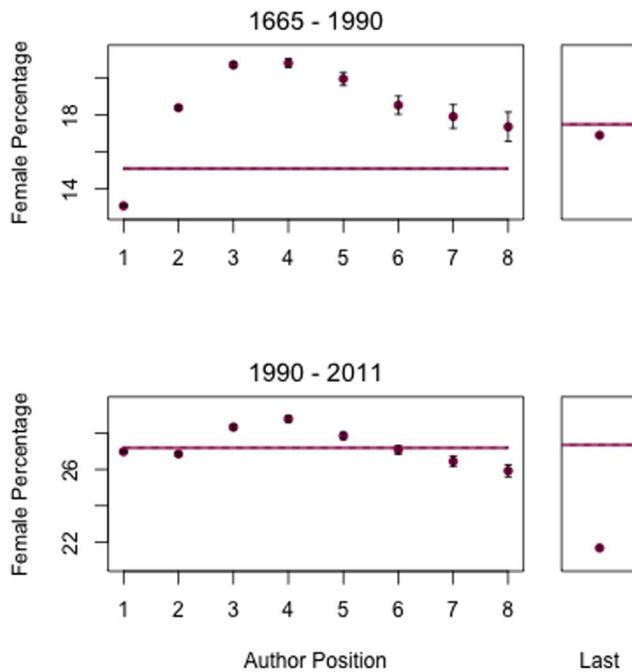doi:10.1371/journal.pone.0066212.t002

Figure 3. Gender as a function of authorship order across the entire JSTOR network dataset. Top panel: 888,060 authorships prior to 1990. Bottom panel: 1,156,354 authorships from 1990 to the present. From 1990 to present, women are no longer severely underrepresented as first author, but they are increasingly underrepresented as last author. Error bars indicate one standard deviation of the binomial distribution. For the graph of author position, the solid line indicates the overall frequency of women in the JSTOR network dataset. For the last-author graph, the point indicates the frequency of women who are last author on papers with at least three authors. The horizontal line in this part of the graph indicates the appropriate comparator: the overall frequency of women in any authorship position on papers with three or more authors.
doi:10.1371/journal.pone.0066212.g003

unbiased, we would expect to see the field-wide gender composition reflected at each author position.

Results

In an interactive online visualization at http://www.eigenfactor.org/gender/, we report the gender composition by authorship position and overall, for each field, subfield, etc., of the JSTOR network dataset. Women represent 21.9% of the gender-identified authorships in the entire JSTOR network dataset, but these authorships are not distributed evenly in time across fields, or across authorship positions. For instance, women represent 17% of total single-authored papers in the JSTOR network dataset, but represented only 12% prior to 1990, while they account for 26% of single-authored papers after 1990. Figure 1 shows that the fraction of female authorships in general has increased substantially since the 1960s. However, some of this increase may result from increased ease of identifying woman authors as individuals become more likely to use first name instead of merely initials.

Studies of the economics literature have noted considerable differences in gender representation in subfields [18,19], and our analysis reveals a comparable pattern across the subfields within the JSTOR network dataset. Even within a field such as sociology that has a relatively even gender balance, different subfields can vary dramatically in gender composition, as illustrated in Figure 2.
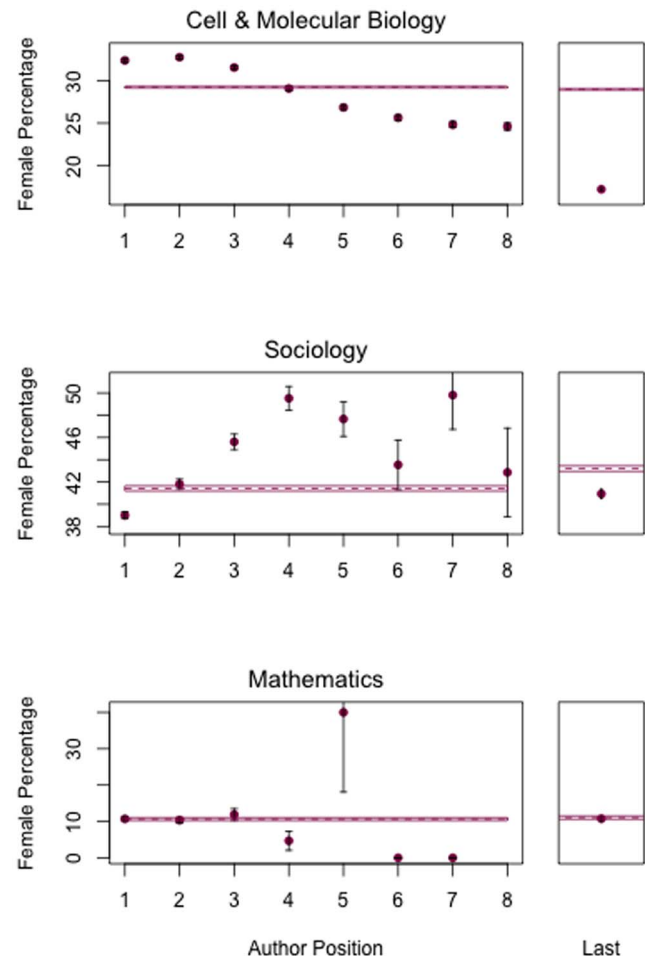

Figure 4. Gender as a function of authorship position in three domains of scholarship from 1990 to present: cell and molecular biology (276,992 authorships), sociology (44,895 authorships), and mathematics (6,134 authorships). In molecular biology, women are overrepresented as first author but underrepresented at the last author position. In sociology, women are underrepresented in both first and last author positions. In mathematics, where the convention is for alphabetical author order [5], women are neither under- nor over-represented at first or last author positions.
doi:10.1371/journal.pone.0066212.g004

Women are not evenly represented across author positions (Table 2). Prior to 1990, women were significantly underrepresented in the first author position; subsequent to 1990 much of this gap has been closed. However, a new gender gap has emerged in the last author position—a position of prestige in the biosciences which represent more than half of the authorships in the JSTOR network dataset (Figure 3). Authorship order patterns vary among fields as well (Figure 4). And because conventions of author order vary across disciplines [20,5], underrepresentation of women in the last author position does not hold up in all fields. In mathematics, for instance, author order tends to be alphabetical irrespective of contribution, and in this field women are evenly represented—albeit at low frequency—across authorship positions.

As expected [21], the proportion of multi-authored papers has increased over time (Figure 5). Some of the pattern in authorship order may be an artifact of this trend in parallel with an increase in the fraction of women over time.
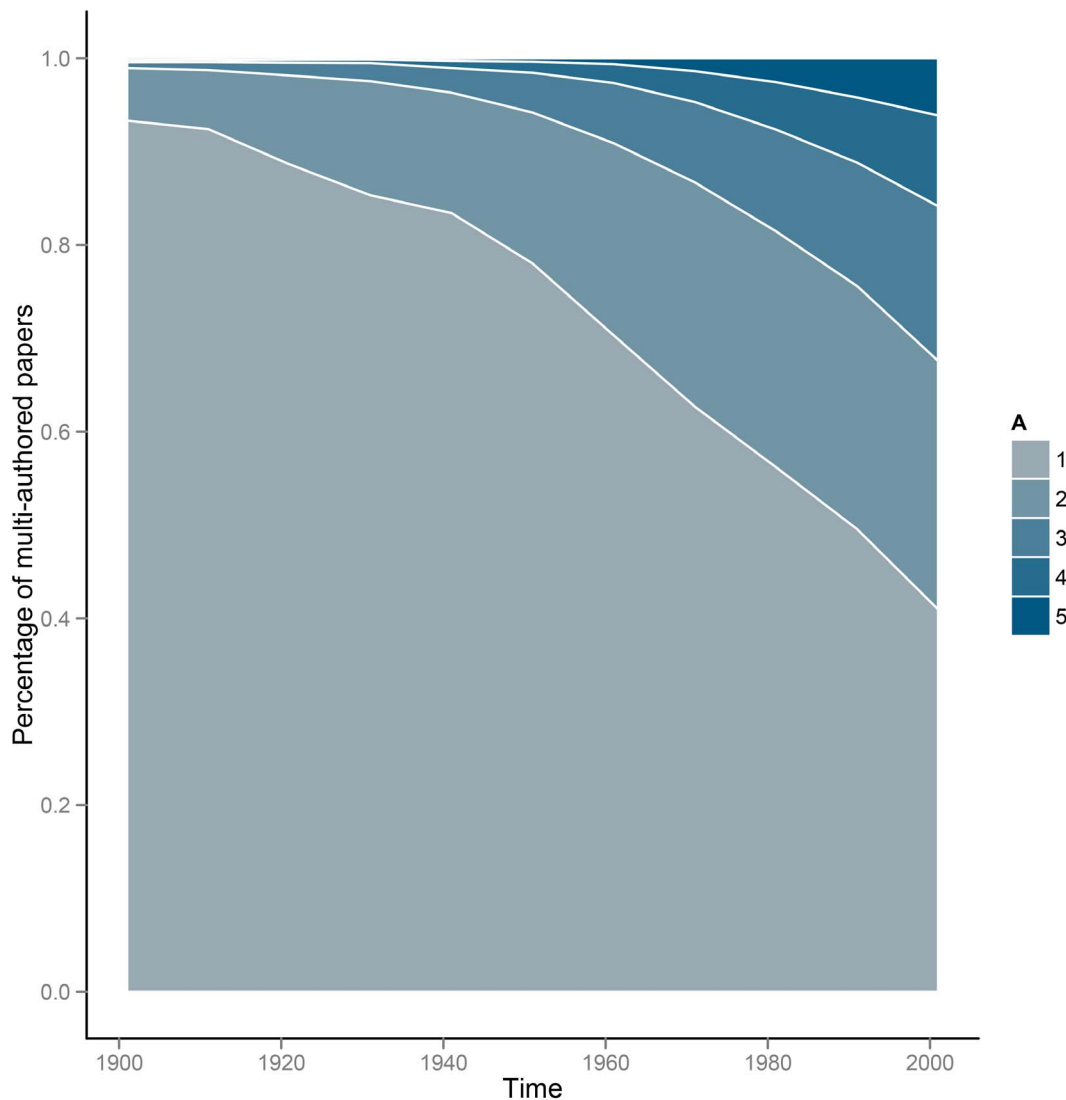
Figure 5. Distribution of author number over time for the JSTOR corpus. Multi-authored papers have increased over time while the fraction of single-authored papers have declined. The y-axis is the percentage of papers with the given number of authors. The legend shows "A", the number of authors on a paper.
doi:10.1371/journal.pone.0066212.g005

Discussion

Only a century ago, women were forbidden from seeking degrees in most universities in Europe [22]. Women seeking a role in academia faced—and continue to face—difficulties at every stage, from admission (Magdalene College at the University of Cambridge was the last all-male college to become mixed, which occurred in 1988), to post-doctoral fellowships [23], to hiring [3], to tenure [24]. As both women and the belief that they belong in universities have infiltrated the academic system, the situation has greatly improved. Women have earned a higher proportion of bachelor's degrees than men since the mid 1980s [25]. In 2004, 48% of PhD recipients were women, up from 16% in 1972 [1]. Despite this increasing equity early in the pipeline, women are still significantly underrepresented in tenure-track and research university faculty positions. Women occupy only 39% of full-time faculty positions and make up an even lower percentage of full professors [1].

Since academic publishing is very important to being hired as a faculty member and being promoted, the under-representation of women as authors in academic publications and in more prestigious authorship positions potentially affects the representation of women faculty in academia. Our research shows that women are increasingly represented in JSTOR network dataset authorships: 27.2% of authorships from 1990–2012 are women compared to just 15.1% from 1665–1989. However, our results also show that the academic publishing environment remains inequitable. For instance, since 1990, women represent only 26% of single-authored papers in the JSTOR dataset.

In many fields, it is not just sheer number of publications, but author order that matters in promotion and tenure decisions. Here we show that women historically have been underrepresented in the first author position, though this is changing, and that women are currently underrepresented in the last author position. (Given these findings, we note the irony of our own authorship order on the present paper.) We should expect some lag between disparity in the first and last author positions, as it takes time for younger

scholars to become leaders of research groups. But the difference between total female authorships and first authorships has been less than 2% since the 1960s, while the discrepancy between total and last authorships remains above 5%. This may reflect a "eaky pipeline" in which women disproportionately leave academia after graduate or postdoctoral training.

While our analysis can clearly delineate gendered patterns in authorship, the data do not allow us to uncover mechanisms that produce the gender disparities we find. Any number of mechanisms could be responsible. One possibility is that women submit fewer papers than men or that their contributions to papers are less significant than their male coauthors, thereby landing them in lower prestige positions on papers. While there is no evidence to support the claim of women's lesser contributions, women are less likely to be involved with collaborative research projects in many scientific fields [26]. A second possibility is that in informal negotiation among a team of authors about author position order, men negotiate more successfully for the more prestigious positions. While we know of no studies that specifically examine authorship negotiations, men, in general, do negotiate more than women [27] and are more likely to self-promote their accomplishments [28]. A third possibility is that there is a bias against women in the review process, such that when they are in the more prestigious author positions, papers of equal quality are less likely to be accepted than when men occupy the prestigious positions. This would produce an underrepresentation of women in journals that do not rely on gender blind reviews. While some have claimed, using correlational data, that gender bias is no longer a factor in producing gender disparities in academia [29], controlled laboratory experiments and field experiments continue to find that biases negatively affect judgments of women [30,31]. For example, a female applicant for science lab manager positions was less likely to be hired than an otherwise identical male applicant, based on judgments of competence by prospective hiring faculty [3]. Furthermore, the report "eyond Bias and Barriers" reviewed the large literature on gender, bias and academic careers and concluded that subtle biases continue to affect women's careers in academia [32].

Our analysis reveals several important patterns: while there have been important gains in parity in the first author position, with the proportion of women in first author positions now even slightly exceeding the overall proportion of female authorships, the proportion of women in the last author position and the proportion authoring overall remain disproportionately low. One strength of this study is that the large dataset represents a significant number of all academics, women and men, across many fields of study and over a large timespan. Though significant progress has been made toward gender equality, important differences in positions of intellectual authorship draw our attention to the subtle ways gender disparities continue to exist. The finding underscores that we cannot yet disregard gender disparity as a notable characteristic of academia.

Author Contributions

Conceived and designed the experiments: JDW JJ MK SJC CTB. Analyzed the data: JW CTB. Wrote the paper: JDW JJ MK SJC CTB.

References

1. West MS, Curtis JW (2006) AAUP faculty gender equity indicators 2006. Technical report, American Association of University Professors.
2. National Research Council (2010) Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty. National Academies Press.
3. Moss-Racusin C, Dovidio J, Brescoll V, Graham M, Handelsman J (2012) Science faculty's subtle gender biases favor male students. Proceedings of the National Academy of Sciences, USA 109: 16474–16479.
4. Conley D, Stadmark J (2012) Gender matters: A call to commission more women writers. Nature 488: 590.
5. Waltman L (2012) An empirical analysis of the use of alphabetical authorship publishing. Journal of Informetrics 6: 700–711.
6. Zuckerman H (1968) Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity. American Journal of Sociology : 276–291.
7. Wren JD, Kozak KZ, Johnson KR, Deakyne SJ, Schilling LM, et al. (2007) The write position. EMBO reports 8: 988–991.
8. Jagsi R, Guancial EA, Worobey CC, Henault LE, Chang Y, et al. (2006) The "gender gap" in authorship of academic medical literature–a 35-year perspective. N Engl J Med 355: 281–7.
9. Feramisco JD, Leitenberger JJ, Redfern SI, Bian A, Xie XJ, et al. (2009) A gender gap in the dermatology literature? Cross-sectional analysis of manuscript authorship trends in dermatology journals during 3 decades. J Am Acad Dermatol 60: 63–9.
10. Sidhu R, Rajashekhar P, Lavin VL, Parry J, Attwood J, et al. (2009) The gender imbalance in academic medicine: A study of female authorship in the united kingdom. J R Soc Med 102: 337–42.
11. Dotson B (2011) Women as authors in the pharmacy literature: 1989–2009. American Journal of Health-System Pharmacists 68: 1736–1739.
12. de Solla Price DJ (1965) Networks of scienti_c papers. Science 149: 510–515.
13. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences, USA 105: 1118–1123.
14. Rosvall M, Axelsson D, Bergstrom CT (2010) The map equation. European Journal of Physics 178: 13–23.
15. Lancichinetti A, Fortunato S (2009) Community detection algorithms: A comparative analysis. Physical Review E 80: 056117:1–11.
16. Rosvall M, Bergstrom CT (2011) Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. PLoS One 6: e18209.
17. Lieberson S, Dumais S, Baumann S (2000) The Instability of Androgynous Names: The Symbolic Maintenance of Gender Boundaries. American Journal of Sociology 105: 1249–1287.
18. Boschini A, Sjögren A (2007) Is team formation gender neutral? Evidence from coauthorship patterns. Journal of Labor Economics 25: 325–365.
19. Dolado JJ, Felgueroso F, Almunia M (2005) Do men and women economists choose the same research fields? Evidence from top 50 departments. Technical report, Centre for Economic Policy Research, London.
20. Endersby JW (1996) Collaborative research in the social sciences: Multiple authorship and publication credit. Social Science Quarterly 77: 375–392.
21. Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. Science 316: 1036–1039.
22. Etzkowitz H, Kemelgor C, Uzzi B (2000) Athena unbound: The advancement of women in science and technology. Cambridge University Press.
23. Wenneras C, Wold A (1997) Nepotism and sexism in peer review. Nature 387: 341–343.
24. Spelke ES, Grace AD (2006) Sex, math, and science. In: Ceci S, Williams W, editors, Why Aren't More Women In Science?: Top Gender Researchers Debate the Evidence., APA Publications.
25. England P, Li S (2006) Desegregation Stalled: The Changing Gender Composition of College Majors, 1971–2002. Gender & Society 20: 657–677.
26. Fox MF (2001) Women, Science, and Academia: Graduate Education and Careers. Gender & Society 1: 654–666.
27. Babcock L, Laschever S (2007) Women Don't Ask: The High Cost of Avoiding Negotiation-and Positive Strategies for Change. New York, NY: Bantam Dell.
28. Rudman LA (1998) Self-Promotion as a Risk Factor for Women: The Costs andBenefits of Counterstereotypical Impression Management. Journal of Personality and Social Psychology 74: 629–45.
29. Ceci SJ, Williams WM (2011) Understanding current causes of women's underrepresentation in science. Proceedings of the National Academy of Sciences USA 108: 3157–3162.
30. Goldin C, Rouse C (2000) Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. American Economic Review 90: 715–741.
31. Correll SJ, Benard S, Paik I (2007) Getting a Job: Is There a Motherhood Penalty? American Journal of Sociology 112: 1297–1339.
32. National Academy of Sciences (2007) Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering. Washington, DC: National Academies Press.
33. Burrelli J (2008) Thirty-three years of women in S&E faculty positions. Infobrief, Science Resources Statistics NSF 08-308, National Science Foundation.