

META-RESEARCH ARTICLE

# Differences in Collaboration Patterns across Discipline, Career Stage, and Gender

Xiao Han T. Zeng<sup>1</sup>, Jordi Duch<sup>1,2</sup>, Marta Sales-Pardo<sup>1,3</sup>, João A. G. Moreira<sup>1</sup>, Filippo Radicchi<sup>4</sup>, Haroldo V. Ribeiro<sup>1,5</sup>, Teresa K. Woodruff<sup>6,7</sup>, Luís A. Nunes Amaral<sup>1,8,9\*</sup>

**1** Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois, United States of America, **2** Department d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain, **3** Department d'Enginyeria Química, Universitat Rovira i Virgili, Tarragona, Spain, **4** Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, Indiana, United States of America, **5** Departamento Física, Universidade Estadual de Maringá, Maringá, Parana, Brazil, **6** Department of Obstetrics & Gynecology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States of America, **7** Institute for Women's Health Research, Northwestern University, Chicago, Illinois, United States of America, **8** Department of Physics & Astronomy, Northwestern University, Evanston, Illinois, United States of America, **9** Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois, United States of America

\* [amaral@northwestern.edu](mailto:amaral@northwestern.edu)



CrossMark  
click for updates

 OPEN ACCESS

**Citation:** Zeng XHT, Duch J, Sales-Pardo M, Moreira JAG, Radicchi F, Ribeiro HV, et al. (2016) Differences in Collaboration Patterns across Discipline, Career Stage, and Gender. *PLoS Biol* 14(11): e1002573. doi:10.1371/journal.pbio.1002573

**Academic Editor:** Lisa Bero, University of California San Francisco, UNITED STATES

**Received:** May 13, 2016

**Accepted:** October 3, 2016

**Published:** November 4, 2016

**Copyright:** © 2016 Zeng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All raw data are available at Figshare: <https://dx.doi.org/10.6084/m9.figshare.3472616>. Underlying data of all figures are available in the Supporting Information.

**Funding:** JD and MSP were funded by the Spanish Ministerio de Economía y Competitividad Award No. FIS2013-47532-C3-1-P (<http://www.mineco.gob.es/>). MSP was funded by the James S. McDonnell Foundation (<https://www.jsmf.org/>). JAGM was funded by Fundação para a Ciência e Tecnologia Grant No. SFRH-BD-76115-2011 (<http://www.fct.pt/>). FR was funded by the National

## Abstract

Collaboration plays an increasingly important role in promoting research productivity and impact. What remains unclear is whether female and male researchers in science, technology, engineering, and mathematical (STEM) disciplines differ in their collaboration propensity. Here, we report on an empirical analysis of the complete publication records of 3,980 faculty members in six STEM disciplines at select U.S. research universities. We find that female faculty have significantly fewer distinct co-authors over their careers than males, but that this difference can be fully accounted for by females' lower publication rate and shorter career lengths. Next, we find that female scientists have a lower probability of repeating previous co-authors than males, an intriguing result because prior research shows that teams involving new collaborations produce work with higher impact. Finally, we find evidence for gender segregation in some sub-disciplines in molecular biology, in particular in genomics where we find female faculty to be clearly under-represented.

## Author Summary

Collaboration plays an increasingly important role in promoting research productivity and impact. What remains unclear is whether female and male researchers differ in their collaboration practices. In our study, we report on an empirical analysis of the complete publication records of 3,980 faculty members in six science, technology, engineering, and mathematical disciplines at select U.S. research universities. First we found that female faculty have significantly fewer distinct co-authors over their careers than males, but that this difference can be fully accounted for by females' lower publication rate and shorter career

Science Foundation Award No. SMA-1446078 (<http://www.nsf.gov/div/index.jsp?div=SMA>). LANA was funded by the Department of Defense's Army Research Office Grant No. W911NF-14-1-0259 ([www.arl.army.mil](http://www.arl.army.mil)) and by the John Templeton Foundation Award No. FP053369-A//39147 (<https://www.templeton.org/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

lengths. Next, we find that female scientists have a lower probability of repeating previous co-authors than males, an intriguing result because prior research shows that teams involving new collaborations produce work with higher impact. Finally, we find evidence for gender segregation in some sub-disciplines in molecular biology, in particular in genomics where we find female faculty to be clearly under-represented.

## Introduction

It is widely acknowledged that collaboration is critical to the scientific enterprise [1–7]. Although the motivations determining collaboration propensity is still the subject of much research, scientists benefit from collaboration both in terms of productivity and impact [8–12]. For example, Bordons et al. [13] showed that for biomedical research there is a positive correlation between productivity and collaboration at the author level, and Wuchty et al. [14] showed that teams produce publications with higher impact than individuals. Moreover, teams that include novel collaborations have a greater likelihood of producing higher impact work [15, 16].

Since research suggests that collaboration patterns affect a researcher's career performance, it is important to understand whether there are gender differences in collaboration patterns [17, 18]. Indeed, Kyvik and Teigen [19] reported that the productivity of both genders is positively correlated with the level of collaboration, and that females have fewer single-author works than males.

Prior research suggests that women tend to be more collaborative and less competitive than men in decision making, making them potentially better collaborators [20–22], but recent studies have reported contradicting results about which gender is more collaborative [23–27].

Because most STEM fields have much larger numbers of males than of females, homophily would suggest that female academics have fewer opportunities for collaboration [28]. McDowell et al. [29] find evidence of gender homophily in collaborator choice among a sample of economists and that females preferentially apply to larger departments to increase their chances of finding collaborators. Bozeman et al. not only find evidence of the same gender homophily [24] but also that, after controlling for gender disparities, females overall collaborate more than males [26].

To investigate the role of gender in collaborative behavior, we perform a large-scale empirical analysis on the publication records of faculty members for six STEM disciplines. Our analyses yield three main findings. First, female faculty have significantly fewer distinct co-authors than male faculty, but that this difference can be fully accounted for by the shorter career lengths of current female faculty and their lower publication rate. Second, female faculty tend to have a lower probability of repeating a collaboration, a strategy that has been shown to produce work of greater impact. Third, for the discipline of molecular biology, we find evidence for gender segregation in some sub-disciplines. In particular, we find that female faculty are clearly under-represented in genomics.

## Data

We obtain complete faculty rosters, as of Fall 2010, for departments of chemical engineering, chemistry, ecology, materials science, molecular biology and psychology from several top research universities in the United States (US) (S1 and S2 Tables). We consider all active faculty members as of 2010, including tenure-track and research faculty, but exclude emeritus professors. We identify the researchers' gender from their departmental website photograph. If they

have no photograph we use their given name to identify the gender (faculty with ambiguous names were excluded). We then obtain bibliometric data for 3,980 faculty members from Thomson Reuters' Web of Science (WoS) based on the biographical information listed on their websites and *curricula vitae*. See [30] for details on data acquisition and validation, and Table 1 for aggregate statistics.

## Results

### Gender differences in number of collaborators

Since scientific publications are the direct product of scientific research and collaboration, the number of distinct co-authors a researcher has accrued throughout her career is a good proxy of how strongly she seeks collaborations. Because collaboration patterns may be discipline-specific, we examine each discipline separately [31]. Moreover, because collaboration patterns may depend on career stage, we also account for career stage in our analyses.

We focus on the number of *distinct* co-authors; that is, we count only once co-authors that appear multiple times in the publications of an individual. We do this because co-authoring publications with new collaborators more likely indicates the introduction of new expertise into the team and the expansion of one's professional network.

We calculate the distribution of total number of distinct co-authors over the career of the scientists in our database. Our raw results show that for all six disciplines, females on average have a significantly lower number of distinct co-authors over their careers than males (Fig 1). However, in order to properly interpret these results, we must account for the fact that until 1980 there were hardly any female faculty, which implies that female faculty typically have shorter career length and thus are likely to have fewer publications than their male colleagues [30]. Moreover, because of the gender gap in the number of publications [30, 32], it is necessary to control for publication rate when comparing the number of co-authors of females and males. Thus, we test the null hypothesis that there is no gender difference in the number of distinct co-authors when controlling for the number of publications (see [Materials and Methods](#)). The confidence intervals constructed under this hypothesis show that once we account for the number of publications, the observed difference in the distribution of the number of distinct co-authors of female and male faculty is not statistically significant (Fig 1).

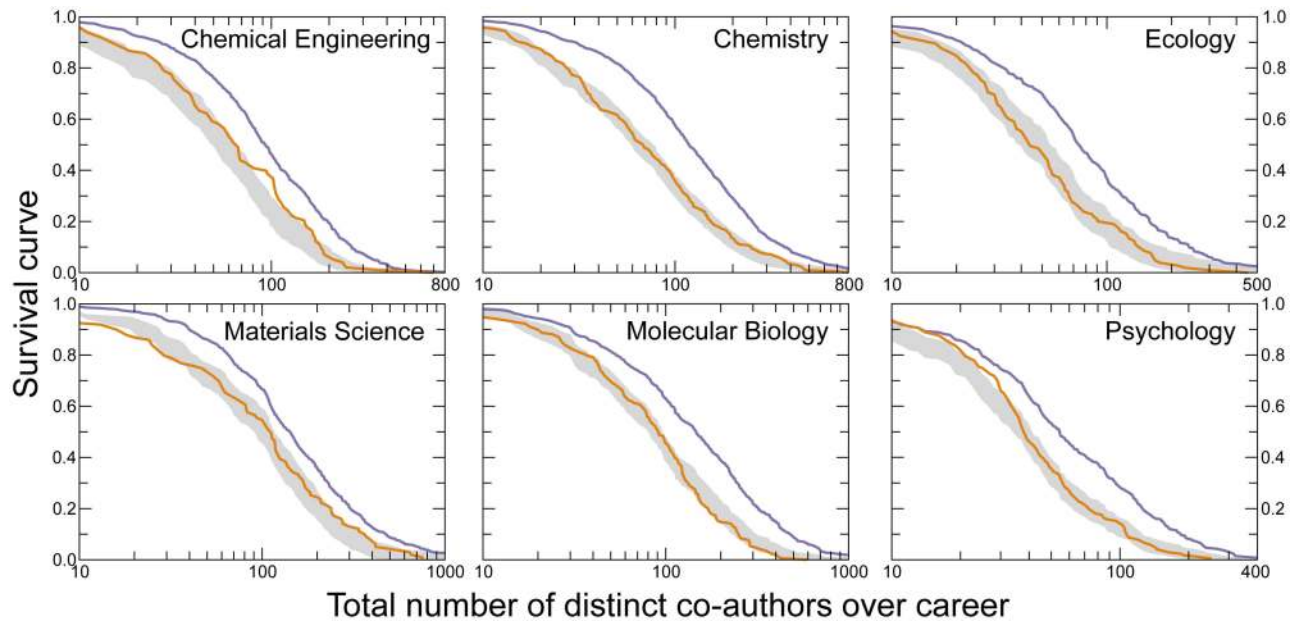
### Repeated co-authors and propensity to collaborate

The data from Fig 1 shows that female and male faculty accrue an average number of new distinct co-authors per publication that is indistinguishable from the average for males. However, this observation does not imply that females and males accrue new collaborators in the same manner, or that they have the same propensity to collaborate.

**Table 1. Characteristics of the faculty cohorts in our study.**

Discipline	Depts.	Faculty			Publications		
		Female	Male	Ratio	Female	Male	Ratio
Chemical Engineering	31	98	567	1:5.8	6,392	66,328	1:10.4
Chemistry	35	198	1,020	1:5.2	13,790	137,723	1:10.0
Ecology	15	106	328	1:3.1	3,976	22,425	1:5.6
Materials Science	26	98	473	1:4.8	9,538	75,373	1:7.9
Molecular Biology	11	168	474	1:2.8	9,882	51,234	1:5.2
Psychology	10	171	279	1:1.6	7,143	20,976	1:2.9
<b>Total</b>	<b>129</b>	<b>839</b>	<b>3,141</b>	<b>1:3.7</b>	<b>50,721</b>	<b>374,059</b>	<b>1:7.4</b>

doi:10.1371/journal.pbio.1002573.t001



**Fig 1. Lower number of publications by female scientists results in lower total number of distinct co-authors.** Survival curve of the total number of co-authors over careers of females (orange) and males (purple). We test the null hypothesis that there is no gender difference in the total number of distinct co-authors for females and males with similar number of publications. The grey shaded region indicates the 95% confidence interval obtained under the null hypothesis. To construct the confidence interval, we generate samples of  $N_F$  males, where  $N_F$  is the number of females in our dataset. For a female with  $n_F$  publications, we select a male whose number of publications falls in the range of  $[0.8 n_F, 1.2 n_F]$  (see [Materials and Methods](#)). Note that the curve for females falls inside the confidence interval, indicating that after correcting for number of publications, females and males have comparable numbers of distinct co-authors over their careers. The curve for males falls outside the confidence interval because some male researchers in the dataset have very large numbers of publications (see Fig 7 of [30]). Data for this figure are in [S1 Data](#).

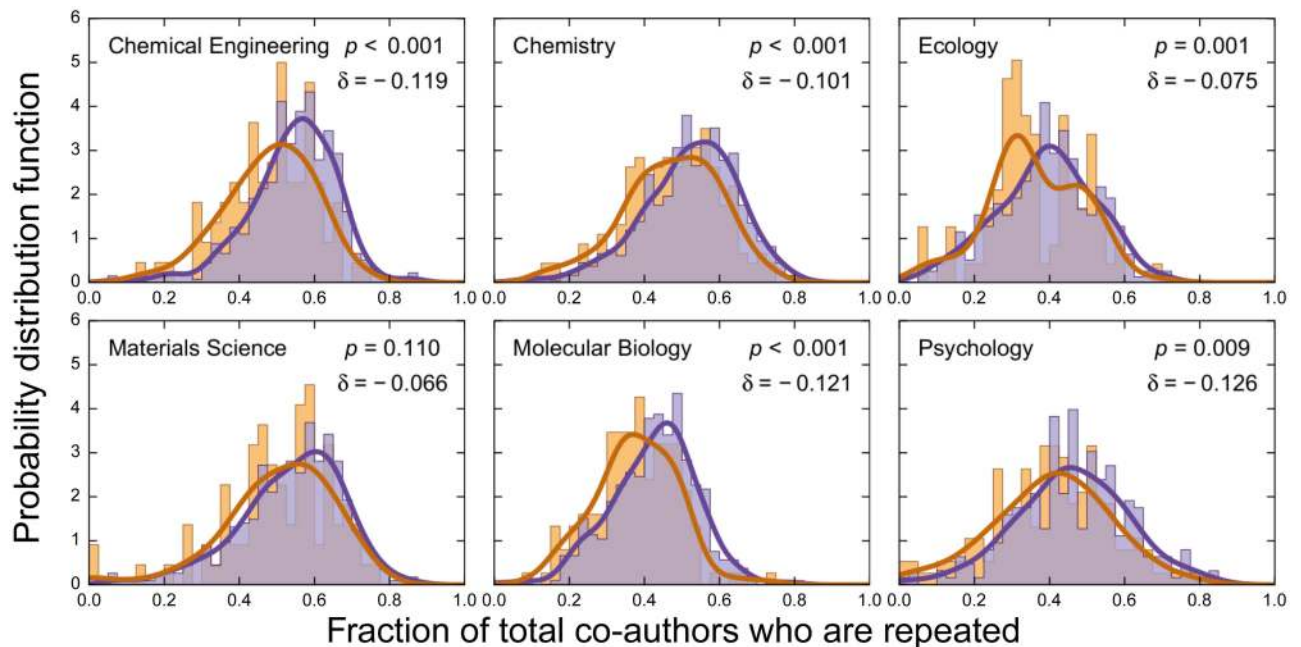
doi:10.1371/journal.pbio.1002573.g001

**Accruing new collaborators.** Consider a publication of researcher  $i$  and  $n_c$  co-authors. The number  $n_n$  of distinct co-authors that  $i$  accrues can be expressed as

$$n_n = n_c (1 - f_r), \tag{1}$$

where  $f_r$  is the fraction of repeated co-authors. [Eq \(1\)](#) makes explicit that both team size (that is,  $n_c$ ) and propensity to repeat collaborations affect the number of new distinct co-authors to be gained from each publication. We first investigate the effect of the repetition of co-authors on the gender disparity in the number of distinct co-authors. Researchers who frequently co-author with the same team will not accumulate co-authors as rapidly as those who seek out new collaboration opportunities. To quantify the tendency to repeat previous co-authors, we calculate  $f_r$  for each author, and obtain the distribution of  $f_r$  for both genders for each discipline. We then test whether the two samples could have been drawn from the same distribution.

We show in [Fig 2](#) the probability distribution functions of  $f_r$  for females and males. The data show that females have an  $f_r$  approximately 20% smaller than males, indicating that female faculty repeat co-authors less frequently than male faculty. More frequent repetition of co-authors may also be an indicator that a few co-authors are responsible for most collaborations. We use the Gini coefficient [33] and the disparity index to quantify the degree of inequality in the distribution of collaboration frequencies, and find that females do tend to distribute their co-authoring opportunities more equally among their collaborators than males ([S1](#), [S2](#) and [S3 Figs](#)). Although the gender difference in the tendency to repeat co-authors is significant, our



**Fig 2. Gender differences in the propensity to co-author with prior collaborators.** Probability distribution of the fraction of total coauthors who are repeated for all females (orange) and males (purple) in the dataset with at least 10 publications. We exclude single-author publications. Orange and purple lines are kernel density estimation of the distributions for females and males with bandwidth given by Scott's Rule [34]. We obtain  $p$ -values for the validity of the null hypothesis that the samples were drawn from the same distribution using the Kolmogorov-Smirnov test. For all disciplines, we find  $\delta = 2(\bar{f}_{r,F} - \bar{f}_{r,M}) / (\bar{f}_{r,F} + \bar{f}_{r,M}) < 0$ , where  $\bar{f}_{r,F}$  and  $\bar{f}_{r,M}$  are the average  $f_r$  of the female and male faculty, respectively. Females have  $f_r$  smaller than those of males, suggesting that, except for materials science, female faculty have a lower propensity than male faculty to repeat collaborations. Data for this figure are in [S2 Data](#).

doi:10.1371/journal.pbio.1002573.g002

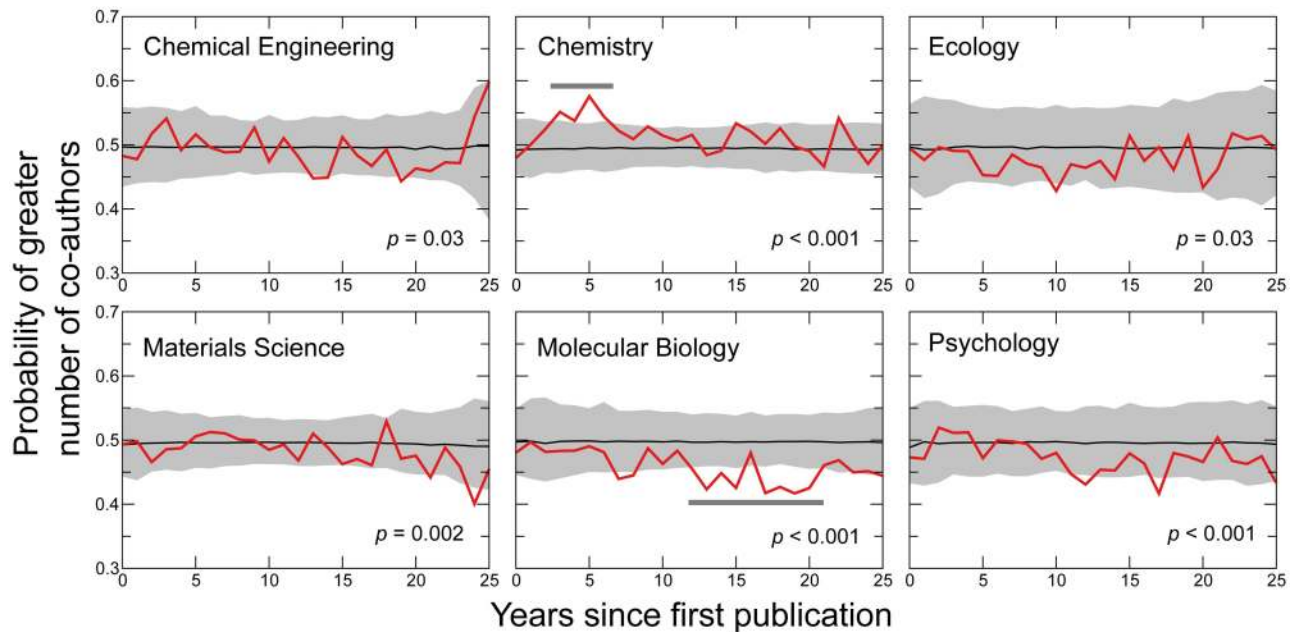
ability to establish its statistical significance on the total number of distinct co-authors is hampered by the heterogeneity in team size and number of publications (S4 Fig).

**Average team size.** We next study the average number of co-authors per publication,  $n_c$ . Researchers who collaborate with larger teams have higher numbers of co-authors per publication. However, the number of co-authors changes as a function of the publication year and author's career stage (S5 Fig). Since female faculty entered academia more recently and on average have shorter career lengths than male faculty [30], we need to account for these two factors when comparing team sizes. In Fig 3 we show that, except for molecular biology, the two genders do not differ significantly in the number of co-authors per publication when their publication year and career stage are taken into consideration.

### The case of molecular biology

Our findings for molecular biology are intriguing. While there are no significant differences during the first ten years, beyond ten years, publications authored by females in molecular biology have significantly lower number of co-authors per publication than those authored by males. To further detail this observation, we bin the publications authored by females according to the number of co-authors, after accounting for increases in team size over the period considered. Assuming that females do not prefer any particular team size, the fraction of publications by females in each bin should remain approximately constant. For each bin, we then calculate how much the observed number of publications by females deviate from the number expected from the null hypothesis using the hypergeometric distribution (see [Materials and Methods](#)).





**Fig 3. Male and female faculty have similar number of co-authors per publication for five other disciplines, but not for molecular biology.** Probability of females having greater number of co-authors per publication in a given year of her career than a male peer at the same career stage (red lines). We use z-scores to account for the increasing size of research teams and the fluctuations over career stage (see [Materials and Methods](#)). We indicate the 99% confidence intervals by the grey areas, and the medians of the probabilities obtained from random ensembles by black lines. The p-values are obtained under the null hypothesis that there is a 99% probability of any value being outside the confidence interval. Note that although the difference in the average size of teams appears to be statistically significant, it is not consistent along the career stage, except for chemistry for the first few years, and for molecular biology in later career stages (dark horizontal bars). Data for this figure are in [S3 Data](#).

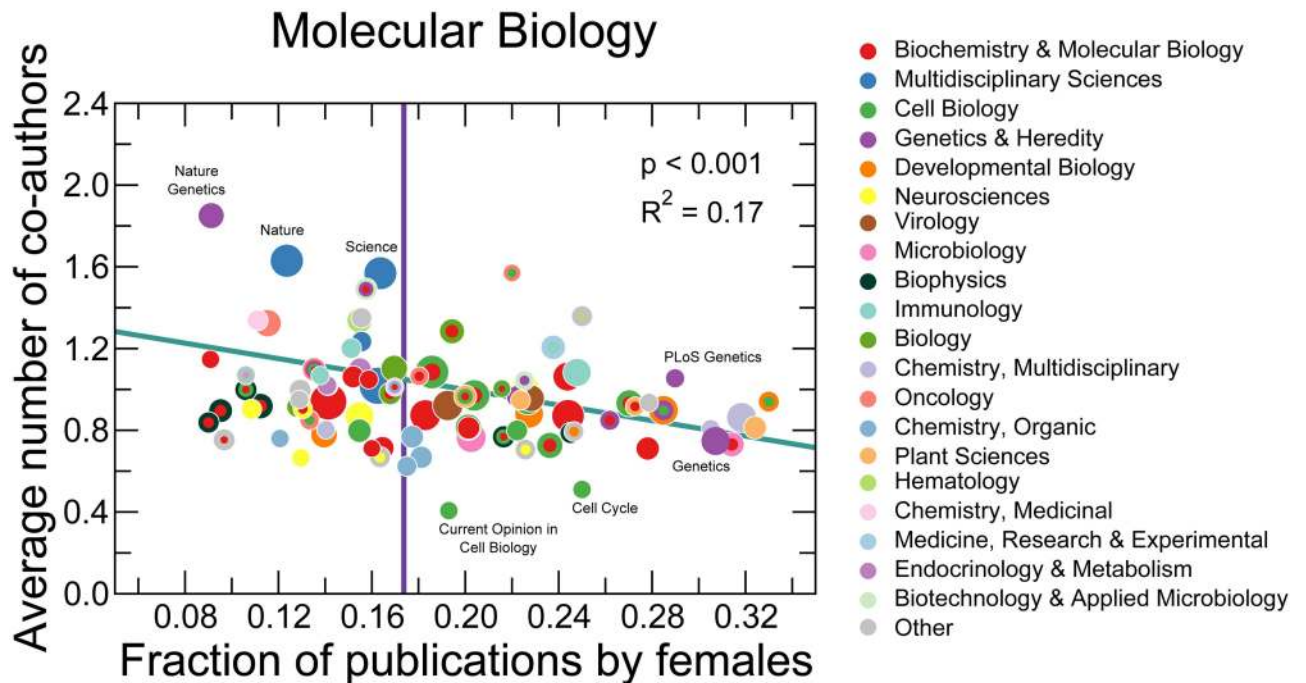
doi:10.1371/journal.pbio.1002573.g003

[S6 Fig](#) demonstrates that female faculty in molecular biology departments have a distinct behavior from females in other disciplines: They consistently author significantly more publications than expected in teams smaller than average, and significantly fewer publications than expected in teams larger than average. We make this fact visually apparent by shading in grey regions where the observed value is significantly different from the null hypothesis.

**Segregation among sub-disciplines.** Although we restrict our analysis to researchers within the same discipline, academic disciplines such as molecular biology comprise several sub-disciplines. If females and males are segregated across sub-disciplines so that more males work in sub-disciplines with large teams, and more females in those with small teams, then this segregation could give rise to the gender gap in the average number of co-authors per publication.

We find that at journal level the average number of co-authors is strongly and significantly anti-correlated with the fraction of publications authored by females ([Fig 4](#)). The strong and statistically significant anti-correlation indicates that females publish more in journals (and, presumably, sub-disciplines) where the typical team size is smaller, and less in those where the typical team size is larger (see [S7 Fig](#) through [S11 Fig](#) for results for other disciplines).

The journal-level analysis strongly suggests the existence of gender segregation across sub-disciplines. However, many journals are multi-topic and even multidisciplinary, thus they may not accurately represent narrower research topics. To overcome this limitation of the journal-level analysis, we must determine the research topic of each publication at a finer scale. To this end, we use a highly accurate and reproducible topic classification algorithm to identify the



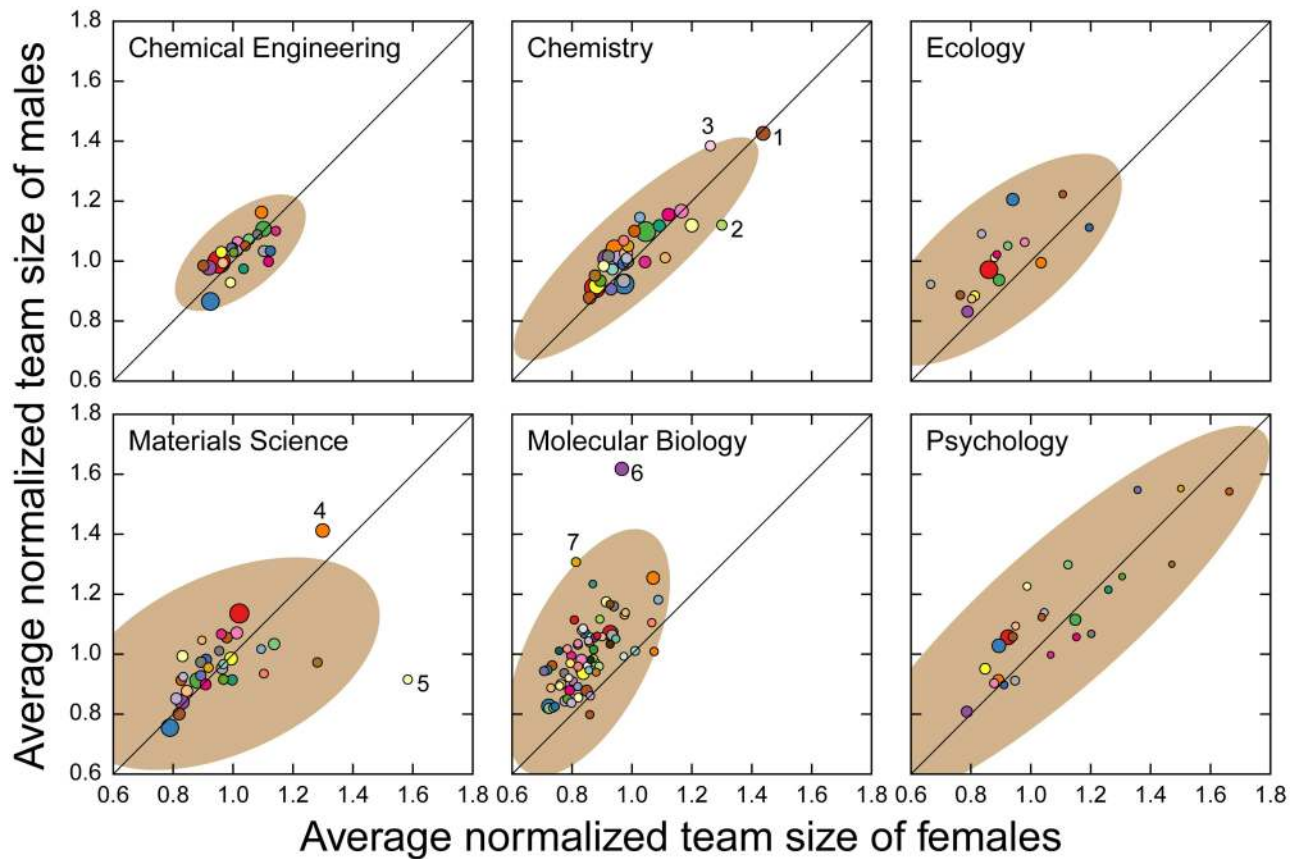
**Fig 4. Female faculty in molecular biology departments publish more in journals and sub-disciplines where typical team size is smaller.** We show correlation between the average number of co-authors corrected for the annual average versus the fraction of publications authored by females, grouped by journal. We only consider publications authored after the tenth year mark in an author's career. We restricted the publication types to "article", "letter", and "note." The size of the circle is proportional to the logarithm of the number of publications in that journal or sub-discipline. We use journal category in the *ISI Journal Citation Report* as the sub-disciplines. Journals with multiple categories are plotted as concentric rings. The purple line indicates the total average fraction of publications by females for all the publications authored by faculty in molecular biology in our cohort,  $f_M$  (17.3%). The blue line is a weighted linear regression, in which we assign to each journal a weight equal to the number of publications. We only include data points within the range of  $[0.5f_M, 2f_M]$ . Data for this figure are in [S4 Data](#).

doi:10.1371/journal.pbio.1002573.g004

topics of publications [35]. We identify a total of 69 topics using the titles and abstracts from the set of 61,116 publications by molecular biology faculty in our database. [S3 Table](#) lists the identified topics and the most representative words and journals associated with them.

For the publications in each topic, we calculate the average team size and fraction of publications by females ([Fig 5](#)). Using a 99% confidence region [36], we identify seven topics that are outliers; of those, two are in molecular biology ([Table 2](#)). All the outlier topics in chemistry and of the outlier topics in materials science actually have larger representations of publications by female faculty and larger team sizes. In contrast, the outlier topics in molecular biology have just larger team sizes. Looking at the representative journals for each of the outlier molecular biology topics, it becomes clear that topic 6 refers to genomics.

Genomics (topic B5) is particularly relevant when attempting to explain the smaller team sizes of female authored molecular biology papers. Genomics is unique because it has a very striking under-representation of females and markedly larger team sizes. Moreover, because it is a topic with a very large number of publications, it strongly affects the characteristics of the entire discipline. These results prompt the question of why females are under-represented in genomics. [S4 Table](#) shows that 19 of the 20 most prolific researchers in our database working in genomics are male. A recent study suggests that the labs of prominent male researchers have lower than average fractions of female graduate students and postdocs [37]. Since the protégés of prominent scientists have such an important role in populating faculty positions in



**Fig 5. Topic dependence of female representation in publications in the six disciplines.** We show the average number of co-authors corrected for the annual average for male faculty versus that for female faculty. Note for molecular biology most of the data points fall above the line  $y = x$ , indicating that for most topics females work in smaller teams than males. We label the seven topics which fall outside the 99% confidence region (brown ellipse) (see Table 2 for topic details). Data for this figure are in S5 Data.

doi:10.1371/journal.pbio.1002573.g005

**Table 2. Topics within considered disciplines that are outliers when considering the differences in average team size between male and female faculty in our database.**

Discipline	Topic	Outlier topic	Representative journals	No. publs.	Normalized ratio by females	Mean normalized team size
Chemistry	C4	1	Cancer Research, Bioconjugate Chemistry, Antimicrobial Agents and Chemotherapy	4,809	1.5	1.43
	C14	2	Nucleic Acids Research, Physical Review E, Genome Biology	1,354	1.3	1.37
	C18	3	Journal of Membrane Science, Radiochimica Acta, Journal of Natural Products	1,399	1.2	1.14
Materials Science	M0	4	Biomaterials, PNAS, Journal of Biological Chemistry	4,547	2.0	1.39
	M29	5	Organometallics, Journal of Chemical Physics, Surface Science	1,742	1.0	0.99
Molecular Biology	B5	6	Nature Genetics, Genetics, Nucleic Acids Research	4,186	1.0	1.51
	B10	7	Molecular Biology and Evolution, Genetics, American Journal of Botany	899	1.1	1.22

**Topic** represents the topic number identified by the topic classification algorithm and is field-specific [35]; **Outlier topic** represents the topic in Fig 5.

doi:10.1371/journal.pbio.1002573.t002



molecular biology, the under-representation of females in those labs propagates all the way to the level of tenured faculty.

In order to investigate the origins of the distinct characteristics of the outlier topics, we turn again to the lists of the scientists with the most publications in each topic (S4 and S5 Tables). We then repeat the analysis of Fig 5 but excluding the publications of the 5 most prolific scientists for each outlier topic. Strikingly, we find that the characteristics of these topics revert to the mean for the entire discipline. That is, the gender of the most prolific authors determines the characteristics of the topic. We believe that this finding raises an important question: Why females have not been able to succeed in genomics in proportion to their numbers? No female in our dataset made it into the top 10 most prolific scientists in genomics, the first female appearing in 12<sup>th</sup> place. If genomics was gender blind, and considering that females comprise 26% of the biology researchers in our database, this would be an unlikely situation ( $p \simeq 0.0095$ ).

## Discussion

A number of recent studies support the hypothesis that there are gender differences in collaboration patterns [17, 18] and that collaboration has a significant impact on scientific productivity and impact [14, 15]. Evidence suggests that self-selection among female researchers due to greater career risks, and female scientists' decreased access to funding can, respectively, cause gender differences in publication rate and impact [29, 30].

Our present analysis conclusively shows that females do have fewer distinct co-authors over their careers, but that this gap can be accounted for by differences in number of publications. We also find evidence for the hypothesis that female scientists are more open to novel collaborations than their male counterparts, a behavior that was shown to correlate with producing work of greater impact [15].

It could be, however, that females have fewer distinct collaborators not purely because, as the females in our cohort they publish fewer publications, but because female scientists do not participate in research teams to the same extent as male scientists. We believe that this possibility is unlikely since there is strong evidence that females are generally more collaborative than males both in academic life [26, 27] and in other realms [20–22].

Concerning our finding that females appear to be more likely to engage new collaborators, it could be that females are simply more effective collaborators and are able to make the most of their lower representation in STEM disciplines. Wolley et al. showed that females typically have greater group intelligence than males [38] giving some credence to this hypothesis. An alternative explanation for the greater repetition of collaborations by males is unwarranted authorship in publications for the purpose of increasing one's publication counts. Anecdotal evidence suggests that, while the number of scientists pursuing such gaming of the system is small, they do tend to be male.

Lastly, our finding of female exclusion from genomics is of particular interest, especially because of what it may imply concerning the cultural milieu of this sub-discipline. The importance of culture on gender segregation is supported by recent studies showing the existence of gender stereotyping in physics and its negative consequences for females in that field [39, 40]. It is known that in some molecular biology sub-disciplines such as telomere research (topic B21) the participation of female scientists has been encouraged. Indeed, 6 of the 10 most prolific researchers in this topic are female (S6 Table). The top three researchers, Elizabeth Blackburn, Virginia Zakian, and Carol Greider conducted their doctoral research under the mentorship of Joseph Gall, who is known for having supported female scientists at a time when misogyny was widely accepted. The important role of prominent scientists in encouraging both males and females to pursue careers in research is also illustrated by William H Bragg's role in

the recruitment of female scientists to crystallography. In contrast, the cultural milieu in institutions such as Genentech [41] likely had a chilling effect on female participation in genomics.

One caveat of our study is that it is limited by the fact that we are only able to track those scientists that persisted within academia. We believe it is important to also investigate to what extent our findings would still hold for scientists that were unable to remain in academic positions at top universities. In a perverse way, it could be that females' propensity to collaborate creates both better publications and a successful research program, and greater risk when the time comes for tenure decisions. Another caveat is that we are not able to identify which coauthors may be trainees (graduate students or post-docs), a situation that in many cases would be more representative of mentorship than of typical collaboration.

## Materials and Methods

### Co-author names matching

To calculate the number of distinct co-authors for a researcher, we used the following procedure. For each researcher, we maintain a set of standardized co-author names. For each co-author name, we convert the name to a string of last name and first name initials. For example, a co-author named "Jane Linda Smith" will be converted to "Smith JL". For each publication, we standardize the names of the co-authors, and add them to the set. We finally count the number of elements in the set.

Note that using this procedure, we treat "Jane Linda Smith" and "Jane Lily Smith" as the same name, because they are both converted to the string "Smith JL". Also, we treat "Jane Linda Smith" and "Jane Smith" as different names, since the former is converted to "Smith JL", while the latter is converted to "Smith J". In reality, for a single author's co-authors, the probability for either case to happen is very small, hence the error rate of our procedure is very low.

### Confidence interval for the survival curve of total number of distinct co-authors

We use matched sampling to obtain the confidence interval for the survival curve of total number of distinct co-authors. We consider the null hypothesis that there is no difference in the total number of co-authors between females and the males with similar number of publications. To construct the confidence interval, we generate samples of  $N_F$  males, where  $N_F$  the number of females in our dataset. For a female with  $n_F$  publications, we select a male whose number of publications falls in the range of  $[0.8 n_F, 1.2 n_F]$ , a range small enough to produce good matches but large enough that there is at least one match. We then compute the survival curve for the obtained sample of male authors. We obtain the confidence interval by repeating this procedure 1,000 times.

The procedure is similar for the null hypothesis that there is no difference in the total number of co-authors between females and the males with equal number of publications, except that the sample of males consists of males who have the same number of publications as the females.

### Measuring gender difference in the distribution of collaboration opportunities

We use two methods, the Gini coefficient and the disparity index, to measure how homogeneously each author distributes all her/his collaboration opportunities among her/his co-authors. A high Gini coefficient or disparity index indicates inhomogeneity of collaboration frequency distribution, where the author collaborates highly frequently with only a small

portion of her/his co-authors, but only a few times with each of the remaining majority. Thus, this author has a high propensity to concentrate her/his collaboration opportunities on a few co-authors. A low Gini coefficient or disparity index indicates that the author collaborates with each of her/his co-authors about equally frequently.

*Gini coefficient.* Consider author  $a$  with  $n_c$  co-authors. For each co-author  $c_i$  of  $a$ , we count the times of collaboration between  $a$  and  $c_i$ ,  $y_i$ . That is, the number of publications  $a$  has co-authored with  $c_i$ . We next arrange  $y_i$  in non-decreasing order, where  $y_i \leq y_{i+1}$ . The Gini coefficient of author  $a$  is calculated as

$$G(a) = \frac{2 \sum_{i=1}^{n_c} iy_i}{n_c \sum_{i=1}^{n_c} y_i} - \frac{n_c + 1}{n_c}. \quad (2)$$

*Disparity index.* We first calculate the weight of collaboration (link) between  $a$  and  $c_i$  as given by Newman [42],

$$w_{ac_i} = \sum_{j=1}^{k_{c_i}} \frac{1}{l_j - 1}, \quad (3)$$

where  $k_{c_i}$  is the number of publications authored by  $a$  and  $c_i$  together, and  $l_j$  is the number of co-authors in publication  $j$ . Then we calculate for  $a$  the summation of the weights of collaboration (strength),

$$s_a = \sum_{i=1}^{n_c} w_{ac_i}. \quad (4)$$

Finally, the disparity index is calculated as

$$\Upsilon(a) = \sum_{i=1}^{n_c} \left( \frac{w_{ac_i}}{s_a} \right)^2 n_c. \quad (5)$$

We obtain the sample of Gini coefficients for female authors,  $\{G_F\}$ , and that for male authors,  $\{G_M\}$ . We then can obtain the significance of the difference between the two samples, by performing a Kolmogorov-Smirnov test on the cumulative distribution function curves of the two samples. We perform the same hypothesis test for  $\{\Upsilon_F\}$  and  $\{\Upsilon_M\}$ .

## Simulating total number of distinct co-authors

We simulate the process of accumulating distinct co-authors and then calculate the total number of distinct co-authors. For each author, we calculate the fraction of repeated co-authors,  $f_r$ . We then generate a list of publications, and record the number of collaborations with each distinct co-author. For each co-author in each publication, we decide if this co-author is a previous co-author with probability  $f_r$ . If yes, we choose a previous co-author with a probability proportional to the times of collaboration with that co-author, and increase the times of collaboration with that co-author by one. Otherwise, we add a new co-author to the list of co-authors. We do not use equal probability when choosing a previous co-author because this would lead to larger number of distinct co-authors than observed.

Initially, we assign to each author 100 publications, in each of which the author has 5 co-authors. The results show that, for most disciplines, females have significantly more distinct co-authors ( $p < 0.0006$ , S4A Fig). This is expected since females repeat co-authors less than

males do. We next introduce the observed heterogeneity in the team size, by keeping the number of publications at 100 while using team sizes sampled from the author's publications. [S4B Fig](#) shows that in this case the gender difference is no longer significant. Finally, we introduce the heterogeneity in the number of publications, by using the actual number of publications and the number of co-authors in each publication ([S4C Fig](#)). Now, females have significantly fewer number of distinct co-authors for most disciplines. These results clearly expose the origins of the results presented in [Fig 1](#) where by controlling for number of publications alone we observed no statistical significant difference between males and females in the number of distinct co-authors.

### Confidence interval for the probability of greater number of co-authors per publication

We consider the probability that publications authored by female authors in our cohort have a larger number of co-authors than publications authored by male authors in our cohort as a function of the career stage of the authors. Since not all the publications are published at the same career stages of the authors, and the size of science teams is increasing with time, we do not consider raw numbers of co-authors but instead standard scores relative to career stages.

Let  $n_i(y)$  denote the number of co-authors of publication  $i$  from discipline  $j$  in year  $y$ , and let  $N_j(y)$  denote the total number of publications published in year  $y$ . We calculate the standard score of publication  $i$  in year  $y$  as

$$z_i(y) = \frac{n_i(y) - \mu_j(y)}{\sigma_j(y)}, \tag{6}$$

where  $\mu_j(y)$  is the average number of co-authors per publication from discipline  $j$  published in year  $y$

$$\mu_j(y) = \frac{\sum_k n_k(y)}{N_j(y)}, \tag{7}$$

$\sigma_j(y)$  is the standard deviation of the number of co-authors per publication published in year  $y$

$$\sigma_j(y) = \sqrt{\frac{1}{N_j(y)} \sum_k [n_k(y) - \mu_j(y)]^2}. \tag{8}$$

We finally consider  $z_i^c(s)$ , the standard score of publication  $i$  as a function of the career stage  $s = y - y_i$ , where  $y_i$  is the year of the first publication of  $i$ 's author. We then calculate for each career stage  $s$  the quantity  $P[z_F^c(s) > z_M^c(s)]$ , representing the probability that a publication authored by a female author has a standard score higher than that of a publication authored by a male author at the same stage of the career as the female author. We also compute the confidence intervals for these probability values, in the null hypothesis that there is no gender difference in the standard scores:

$$H_0 : z_F(t) = z_M(t). \tag{9}$$

We generate the confidence interval valid under this hypothesis using a re-sampling method: The populations of females and males are fixed, the values of all standard scores are also fixed, but values of the standard score are randomly reassigned among publications (this is the same as randomly reassigning the genders to authors). For each random configuration, we



compute again the probability  $P[z_F^c(s) > z_M^c(s)]$  and obtain the confidence interval by repeating this procedure 1,000 times.

### Statistical significance of the number of publications with a given team size

To measure the extent to which females have different team sizes than expected, we use the hypergeometric distribution as the null model. We first account for the increasing trend in the team size over years (S5 Fig). For publication  $i$  with  $n_i$  co-authors from discipline  $j$  in year  $y$ , we calculate the corrected team size,  $v_i(y)$ , by dividing  $n_i$  by the average number of co-authors for all the publications published in year  $y$ ,  $\mu_j(y)$ ,

$$v_i(y) = \frac{n_i(y)}{\mu_j(y)}, \quad \mu_j(y) = \frac{\sum_k n_k(y)}{N_j(y)}, \quad (10)$$

where  $N_j(y)$  is the total number of publications published in year  $y$ . We then bin the publications according to  $v(y)$ .

For the discipline being considered, suppose there are  $N$  publications in total, of which  $N_F$  are authored by females. Consider a bin  $b$  in which there are  $N_b$  publications. If the females collaborate with teams of different sizes with equal probability, then the expected number of publications by females in  $b$  is

$$N_{F,b}^e = N_b \frac{N_F}{N}. \quad (11)$$

Suppose that of the  $N_b$  publications in bin  $b$ ,  $N_{F,b}^o$  are authored by females. The probability of observing  $N_{F,b}^o$  publications by females given by the hypergeometric distribution is then

$$P(X = N_{F,b}^o) = \frac{\binom{N_F}{N_{F,b}^o} \binom{N - N_F}{N_b - N_{F,b}^o}}{\binom{N}{N_b}}. \quad (12)$$

The p-value of observing  $N_{F,b}^o$  is then  $P(X \leq N_{F,b}^o)$ . In S6 Fig we plot  $\log \frac{N_{F,b}^o}{N_{F,b}^e}$  for each bin, and shade the regions where the p-value is significant. We use the Bonferroni correction in which the false discovery rate (FDR) is set to be 0.01. We reject the null model if p-value  $< \frac{0.01}{m}$ , where  $m$  is the number of bins and thus the number of hypotheses.

## Supporting Information

**S1 Fig. Gender differences in the propensity to repeat previous collaboration measured using the Gini coefficient.** Distribution of the Gini coefficient of collaboration heterogeneity [33] for females (orange) and males (purple) in the dataset with at least 10 publications. We exclude single-author publications. We obtain  $p$ -values for the validity of the null hypothesis that the samples were drawn from the same distribution using the Kolmogorov-Smirnov test. For all disciplines, we find  $\delta = 2(\bar{G}_F - \bar{G}_M) / (\bar{G}_F + \bar{G}_M) < 0$ , where  $\bar{G}_F$  and  $\bar{G}_M$  are the average Gini coefficient of the female and male faculty, respectively. Females have Gini coefficients smaller than those of males, suggesting that female faculty have a lower propensity than male faculty to repeat collaborations. Data for this figure are in S6 Data. (EPS)

**S2 Fig. Gender difference in the propensity to repeat previous co-authors measured using the disparity index.** Distribution of the disparity index measuring the repetition of co-authors of females (orange) and males (purple). The  $p$ -values indicate the significance of the gender difference, obtained with Kolmogorov-Smirnov test. The result is in good agreement with that obtained using the Gini coefficient in [S1 Fig](#). Data for this figure are in [S7 Data](#). (EPS)

**S3 Fig. Correlation between Gini coefficient and probability to repeat previous co-authors.** Orange (female) and purple (male) lines are linear fits to data, and  $R_F^2$  and  $R_M^2$  are the corresponding coefficient of determination. Data for this figure are in [S8 Data](#). (EPS)

**S4 Fig. Heterogeneity in the number of publications and team size masks the effect of gender difference in the propensity to repeat co-authors.** Survival curves of the simulated total number of distinct co-authors with fixed number of publications and team size (A), fixed number of publications and team sizes sampled from real data (B), and both number of publications and team sizes from real data (C) for female (orange) and male (purple) faculty in all departments (see [Materials and Methods](#)). We obtained  $p$ -values for the validity of the null hypothesis that the samples were drawn from the same distribution using the Kolmogorov-Smirnov test. Statistical significant results with  $p < 0.01/18 \approx 0.0006$  (Bonferroni correction for multiple hypothesis) are shaded grey. When using fixed number of publications and team size, females have significantly more distinct co-authors. However, the gender difference disappears for most disciplines when using fixed number of publications but real team sizes. When we also use number of publications from the real data, females have significantly fewer distinct co-authors, consistent with [Fig 1](#). Data for this figure are in [S9 Data](#). (EPS)

**S5 Fig. Growth of average number of co-authors during considered period.** Average number of co-authors per publication for females (orange) and males (purple) as a function of publication year. The data are smoothed using a moving averaging method with window size 3. The shaded region indicates the 99% confidence interval obtained with bootstrapping. Data for this figure are in [S10 Data](#). (EPS)

**S6 Fig. In molecular biology departments, female faculty work in smaller teams than male faculty.** Logarithm of the ratio of observed number of publications authored by females over that expected from a hypergeometric distribution (orange circles). The publications are binned by the number of co-authors corrected for the annual average with a bin size of 0.2. The shaded areas indicate that the observed number is significantly different from expected by the model, using the Bonferroni correction by treating each bin as an independent hypothesis test (see [Materials and Methods](#)). The error bars indicate thrice the standard deviation. The black line indicates the ratio of 1.0, and the purple line indicates the average corrected team size. Note that for molecular biology, females have more publications than expected with smaller teams (corrected team size  $< 1.0$ ) and fewer publications than expected with larger teams (corrected team size  $> 1.0$ ). Data for this figure are in [S11 Data](#). (EPS)

**S7 Fig. Correlation between the average number of co-authors corrected for the annual average versus the fraction of publications authored by female faculty in chemical engineering departments.** Publications are grouped by journal. We restricted the publication types to “article”, “letter”, and “note”. The size of the circle is proportional to the logarithm of the

number of publications in that journal or sub-discipline. We use journal category in the *ISI Journal Citation Report* as the sub-disciplines. Journals with multiple categories are plotted as concentric rings. The purple line indicates the total average fraction of publications by females for all the publications authored by faculty in chemical engineering in our cohort,  $f_M$ . The blue line is a weighted linear regression, in which we assign to each journal a weight equal to the number of publications. We only include data points within the range of  $[0.5f_M, 2f_M]$ . Data for this figure are in [S4 Data](#).

(EPS)

**S8 Fig. Correlation between the average number of co-authors corrected for the annual average versus the fraction of publications authored by female faculty in chemistry departments.** See the caption of [S7 Fig](#) for details. Data for this figure are in [S4 Data](#).

(EPS)

**S9 Fig. Correlation between the average number of co-authors corrected for the annual average versus the fraction of publications authored by female faculty in ecology departments.** See the caption of [S7 Fig](#) for details. Data for this figure are in [S4 Data](#).

(EPS)

**S10 Fig. Correlation between the average number of co-authors corrected for the annual average versus the fraction of publications authored by female faculty in materials science departments.** See the caption of [S7 Fig](#) for details. Data for this figure are in [S4 Data](#).

(EPS)

**S11 Fig. Correlation between the average number of co-authors corrected for the annual average versus the fraction of publications authored by female faculty in psychology departments.** See the caption of [S7 Fig](#) for details. Data for this figure are in [S4 Data](#).

(EPS)

**S1 Table. University rankings according to the 2010 edition of the Best Colleges Ranking from US News & World Report [43].** We also show the specialty Graduate School Rankings for Chemical Engineering [44], Chemistry [45], and Ecology [46] when available.

(PDF)

**S2 Table. University rankings according to the 2010 edition of the Best Colleges Ranking from US News & World Report [43].** We also show the specialty Graduate School Rankings for Materials Science [47], Molecular Biology [48], and Psychology [49] when available.

(PDF)

**S3 Table. Research topics in molecular biology.** We show for each topic the list of most representative words and journals. The topic numbers and words are given by the topic classifying method [35], and the journals are those in which the number of publications is significantly more than expected to occur by chance if drawn from a hypergeometric distribution.

(PDF)

**S4 Table. The 20 most prolific scientists in our dataset publishing in topic B5 identified as genomics (outlier topic 6 in [Table 2](#)).**

(PDF)

**S5 Table. The 20 most prolific scientists in our dataset publishing in topic B10 (outlier topic 7 in [Table 2](#)).**

(PDF)

**S6 Table. The 20 most prolific scientists in our dataset publishing in topic B21 identified as telomere research.**

(PDF)

**S1 Data. Data for [Fig 1](#).**

(XLSX)

**S2 Data. Data for [Fig 2](#).**

(XLSX)

**S3 Data. Data for [Fig 3](#).**

(XLSX)

**S4 Data. Data for [Fig 4](#), and [S7 Fig](#) through [S11 Fig](#).**

(XLSX)

**S5 Data. Data for [Fig 5](#).**

(XLSX)

**S6 Data. Data for [S1 Fig](#).**

(XLSX)

**S7 Data. Data for [S2 Fig](#).**

(XLSX)

**S8 Data. Data for [S3 Fig](#).**

(XLSX)

**S9 Data. Data for [S4 Fig](#).**

(XLSX)

**S10 Data. Data for [S5 Fig](#).**

(XLSX)

**S11 Data. Data for [S6 Fig](#).**

(XLSX)

## Acknowledgments

We gratefully thank A. Lancichinetti, D. Mertens, J. Poncela, P. Winter, and members of the SEES Lab for useful discussions and suggestions.

## Author Contributions

**Conceptualization:** XHTZ JD MSP FR HVR TKW LANA.

**Data curation:** XHTZ JD MSP

**Formal analysis:** XHTZ JD JAGM.

**Funding acquisition:** JD MSP JAGM FR LANA.

**Investigation:** XHTZ JD MSP.

**Methodology:** XHTZ JD MSP FR HVR TKW LANA.

**Project administration:** TKW LANA

**Software:** XHTZ JD JAGM.



**Supervision:** JD FR TKW LANA.

**Visualization:** XHTZ JAGM.

**Writing – original draft:** XHTZ JD MSP FR HVR TKW LANA.

**Writing – review & editing:** XHTZ JD MSP JAGM FR HVR TKW LANA.

## References

1. Levine JM, Moreland RL. Collaboration: The social context of theory development. *Personal Soc Psychol Rev.* 2004; doi: [10.1207/s15327957pspr0802\\_10](https://doi.org/10.1207/s15327957pspr0802_10) PMID: [15223516](https://pubmed.ncbi.nlm.nih.gov/15223516/)
2. Stokols D, Hall KL, Taylor BK, Moser RP. The Science of Team Science. Overview of the Field and Introduction to the Supplement. *Am J Prev Med.* 2008; 35(2 SUPPL.):S77–89. doi: [10.1016/j.amepre.2008.05.002](https://doi.org/10.1016/j.amepre.2008.05.002) PMID: [18619407](https://pubmed.ncbi.nlm.nih.gov/18619407/)
3. Falk-Krzesinski HJ, Börner K, Contractor N, Fiore SM, Hall KL, Keyton J, et al. Advancing the science of team science. *Clin Transl Sci.* 2010; 3(5):263–266. doi: [10.1111/j.1752-8062.2010.00223.x](https://doi.org/10.1111/j.1752-8062.2010.00223.x) PMID: [20973925](https://pubmed.ncbi.nlm.nih.gov/20973925/)
4. Cummings JN. Collaborative research across disciplinary and organizational boundaries. *Soc Stud Sci.* 2005; 35(5):703–722. doi: [10.1177/0306312705055535](https://doi.org/10.1177/0306312705055535)
5. Schubert A, Glänzel W. Cross-national preference in co-authorship, references and citations. *Scientometrics.* 2006; 69(2):409–428. doi: [10.1007/s11192-006-0160-7](https://doi.org/10.1007/s11192-006-0160-7)
6. Jones BF, Wuchty S, Uzzi B. Multi-university research teams: Shifting impact, geography, and stratification in science. *Science.* 2008; 322(5905):1259. doi: [10.1126/science.1158357](https://doi.org/10.1126/science.1158357) PMID: [18845711](https://pubmed.ncbi.nlm.nih.gov/18845711/)
7. Milojević S. Principles of scientific research team formation and evolution. *Proc Natl Acad Sci U S A.* 2014; 111(11):3984–3989. doi: [10.1073/pnas.1309723111](https://doi.org/10.1073/pnas.1309723111) PMID: [24591626](https://pubmed.ncbi.nlm.nih.gov/24591626/)
8. Wood DJ. Toward a Comprehensive Theory of Collaboration. *J Appl Behav Sci.* 1991; 27(2):139–162. doi: [10.1177/0021886391272001](https://doi.org/10.1177/0021886391272001)
9. Ahuja G. Collaboration networks, structural holes, and innovation: A longitudinal study. *Adm Sci Q.* 2000; 45(3):425–455. doi: [10.2307/2667105](https://doi.org/10.2307/2667105)
10. Dyer JH. Effective interfirm collaboration: how firms minimize transaction costs and maximize transaction value. *Strateg Manag J.* 2002; 18(7):535–556. doi: [10.1002/\(SICI\)1097-0266\(199708\)18:7%3C535::AID-SMJ885%3E3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%3C535::AID-SMJ885%3E3.0.CO;2-Z)
11. Gajda R. Utilizing Collaboration Theory to Evaluate Strategic Alliances. *Am J Eval.* 2004; 25(1):65–77. doi: [10.1016/j.ameval.2003.11.002](https://doi.org/10.1016/j.ameval.2003.11.002)
12. Uzzi B, Spiro J. Collaboration and creativity: The small world problem. *Am J Sociol.* 2005; 111(2):447–504. doi: [10.1086/432782](https://doi.org/10.1086/432782)
13. Bordons M, Gomez I, Fernandez MT, Zulueta MA, Sndez AM. Local, domestic and international scientific collaboration in biomedical research. *Scientometrics.* 1996; 37(2):279–295. doi: [10.1007/BF02093625](https://doi.org/10.1007/BF02093625)
14. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. *Science.* 2007; 316(5827):1036–1039. doi: [10.1126/science.1136099](https://doi.org/10.1126/science.1136099) PMID: [17431139](https://pubmed.ncbi.nlm.nih.gov/17431139/)
15. Guimerà R, Amaral LAN. Functional cartography of complex metabolic networks. *Nature.* 2005; 433(7028):895. doi: [10.1038/nature03288](https://doi.org/10.1038/nature03288) PMID: [15729348](https://pubmed.ncbi.nlm.nih.gov/15729348/)
16. Katzenback JR, Smith DK. *The Wisdom of Teams.* New York: Harper Business; 2008.
17. Editorial. Science for all. *Nature.* 2013; 495:5. doi: [10.1038/495005a](https://doi.org/10.1038/495005a) PMID: [23472264](https://pubmed.ncbi.nlm.nih.gov/23472264/)
18. West JD, Jacquet J, King MM, Correll SJ, Bergstrom CT. The Role of Gender in Scholarly Authorship. *PLoS One.* 2013; 8(7). doi: [10.1371/journal.pone.0066212](https://doi.org/10.1371/journal.pone.0066212) PMID: [23894278](https://pubmed.ncbi.nlm.nih.gov/23894278/)
19. Kyvik S, Teigen M. Child care, research collaboration, and gender differences in scientific productivity. *Sci Technol Human Values.* 1996; 21(1):54. doi: [10.1177/016224399602100103](https://doi.org/10.1177/016224399602100103)
20. Berdahl JL, Anderson C. Men, Women, and Leadership Centralization in Groups Over Time. *Gr Dyn Theory, Res Pract.* 2005; 9(1):45–57. doi: [10.1037/1089-2699.9.1.45](https://doi.org/10.1037/1089-2699.9.1.45)
21. Bart C, McQueen G. Why women make better directors. *Int J Bus Gov Ethics.* 2013; 8(1):93–99. doi: [10.1504/IJBGE.2013.052743](https://doi.org/10.1504/IJBGE.2013.052743)
22. Kümmerli R, Colliard C, Fiechter N, Petitpierre B, Russier F, Keller L. Human cooperation in social dilemmas: comparing the Snowdrift game with the Prisoner's Dilemma. *Proc R Soc London B Biol Sci.* 2007; 274(1628):2965–2970. doi: [10.1098/rspb.2007.0793](https://doi.org/10.1098/rspb.2007.0793) PMID: [17895227](https://pubmed.ncbi.nlm.nih.gov/17895227/)

23. Cole JR, Zuckerman H. The Productivity Puzzle. In: Maehr ML, Steinkamp MW, editors. *Adv. Motiv. Achiev.* JAI Press; 1984. p.217–258.
24. Bozeman B, Corley E. Scientists' collaboration strategies: implications for scientific and technical human capital. *Res Policy.* 2004; 33(4):599–616. doi: [10.1016/j.respol.2004.01.008](https://doi.org/10.1016/j.respol.2004.01.008)
25. Lee S, Bozeman B. The impact of research collaboration on scientific productivity. *Soc Stud Sci.* 2005; 35(5):673–702. doi: [10.1177/0306312705052359](https://doi.org/10.1177/0306312705052359)
26. Bozeman B, Gaughan M. How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Res Policy.* 2011; 40(10):1393–1402. doi: [10.1016/j.respol.2011.07.002](https://doi.org/10.1016/j.respol.2011.07.002)
27. Abramo G, D'Angelo CA, Murgia G. Gender differences in research collaboration. *J Informetr.* 2013; 7(4):811–822. doi: [10.1016/j.joi.2013.07.002](https://doi.org/10.1016/j.joi.2013.07.002)
28. Kegen NV. Science Networks in Cutting-edge Research Institutions: Gender Homophily and Embeddedness in Formal and Informal Networks. *Procedia—Soc Behav Sci.* 2013; 79:62–81. doi: [10.1016/j.sbspro.2013.05.057](https://doi.org/10.1016/j.sbspro.2013.05.057)
29. McDowell JM, Smith JK. The effect of gender-sorting on propensity to coauthor: Implications for academic promotion. *Econ Inq.* 1992; 30(1):68–82. doi: [10.1111/j.1465-7295.1992.tb01536.x](https://doi.org/10.1111/j.1465-7295.1992.tb01536.x)
30. Duch J, Zeng XHT, Sales-Pardo M, Radicchi F, Otis S, Woodruff TK, et al. The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS One.* 2012; 7(12):e51332. doi: [10.1371/journal.pone.0051332](https://doi.org/10.1371/journal.pone.0051332) PMID: [23251502](https://pubmed.ncbi.nlm.nih.gov/23251502/)
31. Abramo G, D'Angelo CA, Murgia G. The collaboration behaviors of scientists in Italy: A field level analysis. *J Informetr.* 2013; 7(2):442–454. doi: [10.1016/j.joi.2013.01.009](https://doi.org/10.1016/j.joi.2013.01.009)
32. Xie Y, Shauman KA. Sex Differences in Research Productivity: New Evidence about an Old Puzzle. *Am Sociol Rev.* 1998; 63(6):847. doi: [10.2307/2657505](https://doi.org/10.2307/2657505)
33. Ceriani L, Verme P. The origins of the Gini index: extracts from *Variabilità e Mutabilità* (1912) by Corrado Gini. *J Econ Inequal.* 2011; 10(3):421–443. doi: [10.1007/s10888-011-9188-x](https://doi.org/10.1007/s10888-011-9188-x)
34. Scott DW. *Multivariate Density Estimation: Theory, Practice, and Visualization.* New York: John Wiley & Sons; 1992. doi: [10.1002/9780470316849](https://doi.org/10.1002/9780470316849)
35. Lancichinetti A, Siler MI, Wang JX, Acuna D, Körding K, Amaral LAN. High-reproducibility and high-accuracy method for automated topic classification. *Phys Rev X.* 2015; 5:11007. doi: [10.1103/PhysRevX.5.011007](https://doi.org/10.1103/PhysRevX.5.011007)
36. Draper NR, Smith H. *Applied Regression Analysis.* 2nd ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1998. doi: [10.1002/9781118625590](https://doi.org/10.1002/9781118625590)
37. Sheltzer JM, Smith JC. Elite male faculty in the life sciences employ fewer women. *Proc Natl Acad Sci U S A.* 2014; 111(28):10107–10112. doi: [10.1073/pnas.1403334111](https://doi.org/10.1073/pnas.1403334111) PMID: [24982167](https://pubmed.ncbi.nlm.nih.gov/24982167/)
38. Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW. Evidence for a collective intelligence factor in the performance of human groups. *Science.* 2010; 330(2010):686–688. doi: [10.1126/science.1193147](https://doi.org/10.1126/science.1193147) PMID: [20929725](https://pubmed.ncbi.nlm.nih.gov/20929725/)
39. Barthelemy RS, McCormick M. Gender Discrimination in Physics and Astronomy: Graduate Student Experiences of Sexism and Gender Microaggressions. *Phys Rev Phys Educ Res.* 2016;020119(12):1–28. doi: [10.1103/PhysRevPhysEducRes.12.020119](https://doi.org/10.1103/PhysRevPhysEducRes.12.020119)
40. Gonsalves AJ, Danielsson A, Patterson H. Masculinities and experimental practices in physics: The view from three case studies. *Phys Rev Phys Educ Res.* 2016; 12(2):020120. doi: [10.1103/PhysRevPhysEducRes.12.020120](https://doi.org/10.1103/PhysRevPhysEducRes.12.020120)
41. Raab GK. CEO at Genentech, 1990–1995: Oral History Transcript. Bancroft Library, University of California, Berkeley; 2003. Available from: <https://archive.org/details/ceogenentech00raabrich>.
42. Newman MEJ. Coauthorship networks and patterns of scientific collaboration. *Proc Natl Acad Sci U S A.* 2004; 101 Suppl:5200–5205. doi: [10.1073/pnas.0307545100](https://doi.org/10.1073/pnas.0307545100) PMID: [14745042](https://pubmed.ncbi.nlm.nih.gov/14745042/)
43. U.S. News & World Report: Best Colleges Rankings 2010 Edition;. Available from: <https://web.archive.org/web/20100512221918/http://colleges.usnews.rankingsandreviews.com/best-colleges/national-universities-rankings>.
44. U.S. News & World Report: Best Graduate Schools in Chemical Engineering 2011 Edition;. Available from: <https://web.archive.org/web/20100914231426/http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-engineering-schools/chemical-engineering>.
45. U.S. News & World Report: Best Graduate Schools in Chemistry 2010 Edition;. Available from: <https://web.archive.org/web/20091001135906/http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-chemistry-schools/rankings>.

46. U.S. News & World Report: Best Graduate Schools in Ecology 2010 Edition;. Available from: <https://web.archive.org/web/20090428035058/http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-biological-sciences-programs/ecology>.
47. U.S. News & World Report: Best Graduate Schools in Material Engineering 2011 Edition;. Available from: <https://web.archive.org/web/20100915141756/http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-engineering-schools/material-engineering>.
48. U.S. News & World Report: Best Graduate Schools in Molecular Biology 2010 Edition;. Available from: <https://web.archive.org/web/20090428035103/http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-biological-sciences-programs/molecular-biology>.
49. U.S. News & World Report: Best Graduate Schools in Psychology 2010 Edition;. Available from: <https://web.archive.org/web/20100515154410/http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-psychology-schools/rankings>.