报告编号：2022_1090

文献检索报告

根据委托人提供的论文清单，经河海大学图书馆检索，查找卢超的论文被 SSCI 收录的情况，详细收录情况见附件。

检索工具：SSCI（Social Sciences Citation Index 1900-至今）
检索类别：收录
检索时限：1900--2023
检索结果：收录 1 篇，第一作者论文 1 篇，通讯作者论文 0 篇，通讯作者(第一作者除外)论文 0 篇，非第一、通讯作者论文 0 篇

注：
1.检索结果已获得委托人的认可。
2.报告中所陈述的内容均以当时检索到的客观文献为依据，涂改无效，本报告遗失不补。由于报告遗失产生的后果自负！

河海大学科技查新工作站(盖章)

检索人：高依旻

2022 年 06 月 14 日

一、文献类型统计如下

序号	文献类型	第一作者	数量
1	Article	1	1

二、SSCI 检索结果

第 1 条，共 1 条

文献标题: Contributorship in scientific collaborations: The perspective of contribution-based byline orders
作者: Lu, C;Zhang, CW;Xiao, CR;Ding, Y
文献类型: Article
出版物名称:INFORMATION PROCESSING & MANAGEMENT **出版年:**MAY 2022 **卷:**59 **期:**3 **页数:**14 **DOI:**10.1016/j.ipm.2022.102944
Web of Science 核心刊的"被引频次":0
被引频次合计:0
入藏号:WOS:000803814200004
作者地址:[Lu, Chao; Xiao, Chengrui] Hohai Univ, Business Sch, Nanjing, Peoples R China. [Zhang, Chenwei] Univ Hong Kong, Fac Educ, Hong Kong, Peoples R China. [Ding, Ying] Univ Texas Austin, Sch Informat Sci, Austin, TX USA.
通讯作者地址: Zhang, CW (通讯作者)，Univ Hong Kong, Fac Educ, Hong Kong, Peoples R China.
电子邮件地址:chwzhang@hku.hk
语种: English
Web of Science 类别: Computer Science, Information Systems; Information Science & Library Science
学科类别: Computer Science; Information Science & Library Science
基金资助机构和授权号

基金资助机构	授权号
National Natural Science Foundation of China	**B220201058**
Fundamental Research Funds for the Central Universities	**72004054**

基金资助正文: This article is an outcome of the youth project "Study of Scientific Collaborators? Scientific Effectiveness from The Perspective of Division of Labor" (ID: 72004054) supported by the National Natural Science Foundation of China and the project "The Causal Effect of Team Diversity on Team Performance" supported by the Fundamental Research Funds for the Central Universities (ID: B220201058) . The authors appreciate the anonymous reviewers for their detailed comments for the improvement of this work.
作者全名: Lu, Chao;Zhang, Chenwei;Xiao, Chengrui;Ding, Ying
通讯作者: Zhang, CW
页数:14

Information Processing and Management 59 (2022) 102944

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Contributorship in scientific collaborations: The perspective of contribution-based byline orders

Chao Lu [a], Chenwei Zhang [b,*], Chengrui Xiao [a], Ying Ding [c]

[a] *Business School, Hohai University, Nanjing, China*
[b] *Faculty of Education, The University of Hong Kong, Hong Kong, China*
[c] *School of Information Science, The University of Texas at Austin, TX, USA*

ARTICLE INFO

Keywords:
Contributorship
Co-contributorship network
Author byline orders
Author contribution statement

ABSTRACT

Scientific collaboration empowers scholars to build larger teams and produce more high-quality knowledge. However, with insufficient microscopic examination of scientific collaboration, i.e., the interactions between collaborators, little is currently known about whether the contributing roles of collaborators are fairly and accurately represented on the bylines of scientific papers. To fill this gap, the current study examines how the different roles of collaborators are connected to the order of their names in article bylines across disciplines and team sizes. A dataset of 105,192 articles containing author contribution statements was compiled and analyzed to investigate the byline order distributions of three different contributing roles: versatiles, specialists, and team-players. We discovered that, across all disciplines, the order of names in article bylines usually represented authors' contributions. Versatiles were more likely to be first authors in the byline, followed by teamplayers and specialists. We also found that the division of labor in larger teams varied between disciplines. In some subjects, the three contributing roles disappeared as the size of teams increased, while in others, they remained distinct. Finally, larger team sizes were associated with a weaker relationship between byline ordering and contributing roles. These findings advance studies of scientific collaboration and enrich the literature on the evaluation of scientific performance.

1. Introduction

Collaboration is vital to the production of scientific knowledge (Larivière et al., 2015; Lee et al., 2019; Wuchty et al., 2007). It enables researchers to build larger teams (Cronin, 2001; Chawla, 2019) and produce more high-quality knowledge (Amjad et al., 2017; Larivière et al., 2015). Scholars from bibliometrics and related fields have investigated the mechanism of scientific collaboration (e.g., Lee et al., 2019; Wu et al., 2019) by examining authorship, i.e., the byline information provided for publications (e.g., Bu et al., 2020; Perneger et al., 2017). This research has developed our understanding of the structure of scientific collaborations (e.g., Arroyo Moliner et al., 2017), their dynamics (e.g., Lee et al., 2019), and interactions between scholars from an external/macro perspective (Zhao et al., 2021).

However, the macro perspective through authorship on scientific collaborations often overlooks the specific interactions between collaborators on a particular project, i.e., the specialization and division of labor within it. To capture the fine-grained detail of

* Corresponding author.
E-mail address: chwzhang@hku.hk (C. Zhang).

https://doi.org/10.1016/j.ipm.2022.102944
Received 16 November 2021; Received in revised form 6 April 2022; Accepted 10 April 2022

scientific collaboration demands close attention to the interactions between actual collaborators (Haeussler & Sauermann, 2020; Lu et al., 2020; Sauermann & Haeussler, 2017). Observational studies of the micro-world of scientific collaborations are relatively rare in the literature (e.g., Lu et al., 2020; Ni et al., 2021) but have found that collaborators usually form small groups to work together as co-contributors (described as *teamplayers* by Lu et al. (2020)), or work independently as individual contributors (*specialists*), or even switch between the two (*versatiles*) to achieve their common research objectives (Lu et al., 2020). It is often difficult to assess the contribution of each collaborator fairly, especially when collaborations are made on both an individual and co-contributory basis (Bao & Zhai, 2017; Xu et al., 2016). This is problematic because the inaccurate assessment of collaborators' contributions may harm the sustainability of scientific collaborations (Wang et al., 2020). Given that contribution-based byline order is usually the primary source for deciding collaborators' credit for contribution, it is valuable to examine whether the contributorship, especially the co-contributorship, can be well encapsulated by the byline order; otherwise, it would be difficult for us to fairly assess each collaborator's contribution to their scientific publications.

To that end, the present study explored how the different contributing roles of collaborators[1] (as states of contributorship) were related to byline order across disciplines and team sizes. We examined whether versatiles, specialists, and teamplayers were more likely to feature at the start of the byline or at the end, and whether this varied according to the size and disciplinary area of teams. To this end, we compiled and analyzed a dataset of 105,192 research articles containing author contribution statements and byline order information. This data was obtained from 173,472 full-text articles published in a range of Public Library of Science (PLOS) journals on different subjects, which we had collected in our previous two studies (Lu et al., 2019a; Lu et al., 2020).

The rest of this article is structured as follows. Section 2 provides an overall review of the literature. Then, Section 3 describes our methodology, introducing our research questions, our data collection and the main methods applied in our study. Section 4 then analyses our main findings to answer our four research questions. Finally, Section 5 summarizes our main results and presents our conclusions.

2. Literature review

In this section, we first introduce studies of scientific collaboration from both the macroscopic and microscopic perspectives. We then review research into author contribution assessment that examines byline order information.

2.1. Macro-level research on scientific collaboration

The presence of two or more authors on the byline of a publication represents a scientific collaboration (Gazni et al., 2012). Byline analyses in bibliometric studies of collaboration have mainly focused on the perspectives of author, institution, country, and academic discipline. Gazni et al. (2012) analyzed 13,917,488 publications harvested from the Web of Science, finding that team size varied considerably according to the field of study, and that larger research teams tended to be more diverse, incorporating collaborators from multiple institutions and countries. Bu et al. (2018) developed a new bibliometric indicator of the persistency of scientific collaborations and found that persistent transdisciplinary collaborations were more likely to produce high-impact research. On the other hand, adopting the method proposed by Uzzi et al. (2013), Wagner et al. (2019) investigated novelty and conventionality in international collaborations and found that the research carried out by such collaborations was usually of a conventional nature.

Other bibliometric studies of collaborations have considered their effect on scientific activity. Collaborative work is associated with increased numbers of citations (Nielsen & Andersen, 2021; Zhang et al., 2020). By using paired matching analysis, Li et al. (2019) determined that junior scholars who collaborated with high-achieving colleagues in their early careers increased their impact and were also more likely to become elite scholars. Comparing the collaborative behavior of scientists who left academia earlier than expected with those who maintained their careers over the long term, Zhao et al. (2021) found that scientists who leave scientific work earlier than expected tended to work in larger teams but were less likely to collaborate with highly cited or senior scholars early in their careers, underlining the importance of such collaborations.

Recent bibliometric studies of scientific collaboration have taken a range of approaches to social network analysis. These aim to uncover the underlying structures of scientific collaboration networks, capture network dynamics, and reconstruct actual collaboration networks (e.g., Wang & Barabási, 2021). Such analyses treat scholars as nodes in a collaboration network, with two nodes considered connected if two authors have published at least one article together. Guimera et al. (2005) studied the assembly mechanism of collaboration teams over time in multiple fields and found that team size, the recruitment of newcomers, and incumbents' tendency to repeat prior collaborations were factors in the self-assembly of teams as well as the success of their collaborations. Feng and Kirkley (2020) investigated collaborative preferences at the individual, pair, and team levels in a collaboration network in education. They found that interdisciplinarity was conferred on projects via the diverse research experiences of individual scholars, rather than by diversity at the pair or team levels.

The macro perspective has advanced understanding of various aspects of scientific collaboration, such as external collaboration patterns, the evolution of collaboration over time, and the topology and dynamics of collaboration networks. However, such research has usually overlooked the specific interactions between collaborators on a team or a certain project, i.e., the specialization and division of labor within them. Understanding these interactions demands careful examination of the internal dynamics among

[1] In this manuscript, we used "collaborator" instead of "coauthor" to emphasize the collaborating relationship between coauthors in a project.

collaborators within a particular team.

2.2. Micro-level research on scientific collaboration

Alongside the macro view, organizational and human behavior researchers have investigated the interactions between collaborators using primarily qualitative methods, such as questionnaires, interviews, and field studies (e.g., Chen et al., 2021; Chung et al., 2016; Delfgaauw et al., 2018). While the sample sizes used in such studies are limited, their findings can inform and gain further validation via large-scale research into scientific collaboration.

Recently, increasing research attention has been paid to author contribution statements (e.g., Corrêa et al., 2017; Haeussler & Sauermann, 2020; Larivière et al., 2016; Lu et al., 2020; Sauermann & Haeussler, 2017; Yang et al., 2017). This new form of data discloses the contributions made by each collaborator on a project, thus illuminating their roles and interactions concerning the division of labor and specialization. Initially, Larivière et al. (2016) and Sauermann and Haeussler (2017) introduced this new research topic by analyzing the distributions of the five main author tasks in articles published in PLOS family journals. They further investigated the relationship between author contribution and citation count, academic age, discipline background, and team size. Yang et al. (2017) analyzed the relationships between byline order and author contribution statements in three medical journals using descriptive statistics. Corrêa et al. (2017) subsequently introduced bipartite network modeling of author–task pairs from the statement data to assess author contributions. Inspired by this study, Lu et al. (2020) created a new network model—termed the *co-contributorship network*—to understand scientific collaboration and the division of labor therein, confirming that the division of labor is fundamental to scientific collaboration and is more prominent in larger teams. They also described three distinct contributing roles—*versatiles, specialists*, and *teamplayers*—based on their roles in the network and contribution to the teams. These roles are defined in the following subsection. Employing an alternative methodology, Haeussler and Sauermann (2020) confirmed Lu et al.'s (2020) observations. Furthermore, their research demonstrated that the division of labor expanded in interdisciplinary teams.

2.3. Byline information in author contribution assessment

The order of contributors' names in article bylines is a critical way to demonstrate accountability and allocate credit for the work that has been done. It must be accurate in order not to jeopardize the research team's sustainability (Teixeira da Silva, 2021; Wang et al., 2020). The order of authors in the byline can be arranged in several ways, most frequently based on contributions or the alphabetical order of their names (Ali, 2021; Bavdekar, 2012; Teixeira da Silva, 2021). In contribution-based byline ordering, it is assumed that the first author made the largest contribution to the study, including drafting the article and conducting experiments and analyses (Bhandari et al., 2014; Perneger et al., 2017; Yang et al., 2017). Corresponding authors, who are usually—but not necessarily—the first or last on the byline (Mattsson et al., 2011; Teixeira da Silva, 2021) are more active in areas such as obtaining funding, revising the article, designing the study, and providing support (Lu et al., 2019b, 2020; Perneger et al., 2017). Additional authors are believed to make smaller contributions (Corrêa et al., 2017; Perneger et al., 2017). Thus, most studies view contribution-based byline ordering as representing each author's contribution to the article and, therefore, an important means of allocating credit to authors (e.g., Bao & Zhai, 2017; Kim & Kim, 2015; Rahman et al., 2017; Xu et al., 2021).

3. Methodology

In this section, we provide a brief description of our methodology. First, we propose our research questions, then describe our dataset, and, finally, detail the specific methods employed to analyze the data.

3.1. Research questions

Based on the above review of literature, we formulated four questions to answer in this study. The last two questions reflect our interest in how subject and team size are related to contributorship in scientific collaborations and its relationship with byline order distributions.

RQ1. Are collaborators' contributions reflected by their position in the byline?

In many subjects, contribution-based author ordering is a convention followed when adding the authors' names to the byline page before submission. Intuitively, one author's contribution should be faithfully reflected by his/her position in the byline. However, in practice, this rule may be inconsistently applied and potentially open to manipulation, gender bias, and unjust allocation of credit (Bavdekar, 2012). Author contribution disclosure statements are intended to address this issue (Frische, 2012), by defining collaborators not merely as co-authors but in terms of their specific contributions. Thus, the first step in our research was to examine whether byline orders accurately reflect the actual weight of the contributions made by collaborators.

RQ2. Do the three contributing roles (versatiles, specialists, and teamplayers) maintain the same or different byline order distributions?

The contribution statement not only discloses each collaborator's contribution but uncovers the division of labor in the collaboration, indicating overall contributorship on the project. It is thus ideal for a data-driven approach to understanding the mechanism of scientific collaboration. Our recent study identified three types of contributorship (or contributing roles) across subjects and in various team sizes via modeling of the co-contributorship network (Lu et al., 2020). The three roles of *versatiles, specialists*, and *teamplayers* contribute differently to collaborations and interact in diverse ways. *Versatiles* are those who work both interdependently and independently on tasks that often require leadership and supervision. *Teamplayers* are those who only work collectively on more general scientific tasks such as data analysis and performing experiments. *Specialists* are those who only work independently, either on leadership tasks, e.g., as supervisors or principal investigators, or make marginal contributions such as enrolling participants. By comparing the distributions of these roles on the byline, different forms of contributorship can be observed. For instance, as the most likely leaders of a study, are versatiles positioned at the start of bylines? Are specialists, who fulfill the contrasting role, randomly distributed among other authors? Are teamplayers, whose contributions are limited to general tasks and non- leadership roles, placed at the end of the bylines?

RQ3. Do the byline order distributions of the contributing roles vary by discipline?

Contribution-based authorship ordering practice varies according to the domain of study (Burrows & Moore, 2011). For example, in biomedical research, the last author can be a senior author or someone who coordinates the project, whereas in physics, the largest contribution is made by the first author while the rest may be placed alphabetically on the byline. It is thus possible that the subject area might affect the distribution of the three contributing roles across the byline.

RQ4. Does team size affect the division of labor and the byline ordering in collaborations within the same discipline?

Team sizes tend to vary by subject (Larivière et al., 2015). Therefore, within each subject area, the size of teams might also be related to the division of labor on the research and the byline order of the corresponding paper (Haeussler & Sauermann, 2020; Larivière et al., 2015; Lu et al., 2020).

3.2. Data collection

To answer our research questions, 173,472 full-text articles published in *PLOS* between 2006 and 2015 were collected in XML format, together with their metadata, which included author information (see Fig. 1). First, author contribution statements from the 161,671 research articles in this corpus were extracted and parsed using natural language processing techniques assisted by manual correction when necessary (see the parsed sample in Table 1). A more detailed account of the processing procedure can be found in our previous study (Lu et al., 2020, pp. 1165–1166). We retained only those statements that had been correctly parsed in full, leaving author-contribution pairs from 138,787 research articles.

After parsing the data, we matched the authors' full names in the byline with their abbreviations provided in the author contribution statement (see Fig. 2). In the absence of hyperlinks between the abbreviations and the full names from the original XML files, we used string matching to link the initial letters of the full names in the bylines to the abbreviations in the contribution statements, an effective strategy in most cases. Similar to the parsing process, we kept only articles for which this process had been successfully completed. The final dataset containing the authors' contribution information and the byline order included 105,192 articles obtained from *PLOS* journals.

3.3. Methods

3.3.1. Network construction

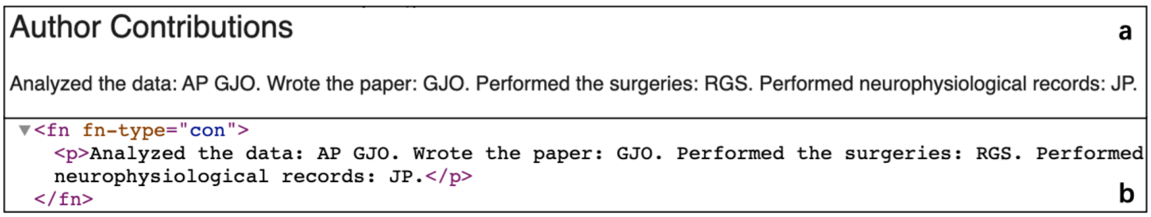
We built a co-contributorship network for each scientific collaboration by performing a one-mode projection of the author-task bipartite network onto the parsed author-task pairs. We also created self-edges for a node if a task was performed by only one person. In a co-contributorship network, each node can be identified as one of the three states of contributorship (or contributing roles), i.e., *versatiles, specialists*, and *teamplayers*. Teamplayers have no self-edges because they perform all their tasks collaboratively; specialists only have self-edges (s) (e.g., C_1) because they work fully independently, while versatiles are those who have both self-edges and normal edges (e.g., C_2; Lu et al., 2020, p. 1167).

3.3.2. Normalized byline order

Byline order is the order in which authors' names are assigned to their publications, usually representing the relative weight of their contributions to the team's work (Corrêa et al., 2017). We used Formula (1) to calculate each author's normalized byline order in every single article, mitigating the effect of different numbers of authors among the articles.

$$NB_i^j = \frac{B_i^j - 1}{N_i - 1},\ N_i \geq 2,\ 0 \leq NB_i^j \leq 1 \tag{1}$$

In this formula, B_i^j is an author's byline order in his/her collaborated article i and N_i denotes the total number of authors in article i and NB_i^j represents the normalized byline order of the author, which is in the range of [0,1].



Author Contributions a

Analyzed the data: AP GJO. Wrote the paper: GJO. Performed the surgeries: RGS. Performed neurophysiological records: JP.

```
▼<fn fn-type="con">
  <p>Analyzed the data: AP GJO. Wrote the paper: GJO. Performed the surgeries: RGS. Performed
  neurophysiological records: JP.</p>
</fn>
```
b

Fig. 1. An author contribution statement presented on the HTML page and marked in the XML file.

Table 1
Author-contribution pairs completely parsed from a piece of contribution statement.

Id	Authors	Task
1	AP; GJO	Analyzed the data
2	GJO	Wrote the paper
3	RGS	Performed the surgeries
4	JP	Performed neurophysiological records

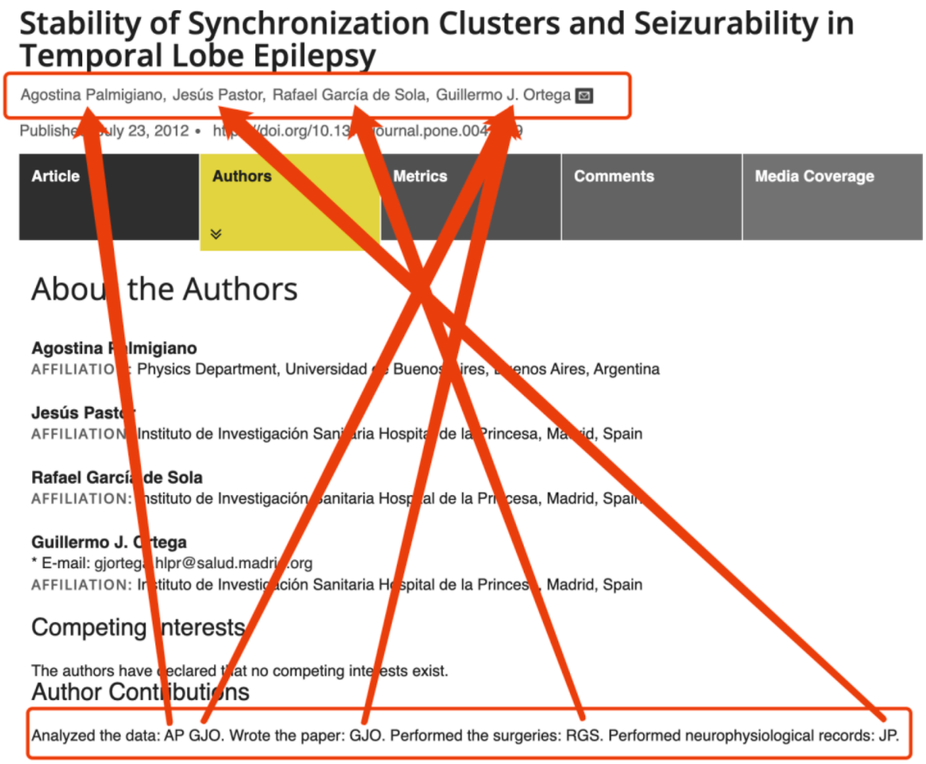


Fig. 2. Matching authors' full names with their abbreviations indicated in the author contribution statement.

3.3.3. *Author grouping by normalized byline order*

To identify the relationship between collaborators' byline orders and their contribution statements, we allocated each collaborator to one of five tiers according to their normalized byline orders, as shown in Table 2.

3.3.4. *Task frequency as a proxy for contribution assessment*

To identify the contribution made by each author to the collaboration, we followed Corrêa et al. (2017) by comparing the frequency with which each task was performed by different tiers of authors. We recognized that not all submissions strictly followed the PLOS declaration rules for author contribution statements, which resulted in varied descriptions for certain tasks, e.g., "wrote the paper" vs. "wrote the manuscript" (Lu et al., 2020). In these cases, we manually annotated the data to ensure uniformity.

3.3.5. *Disciplinary distribution*

Only 89,149 of our 105,192 articles were identifiable from the Web of Science records using the NSF subject classification system (Lu et al., 2019b) as shown in Table 3.

3.3.6. *Frequency weighting resampling*

While our dataset was multidisciplinary, particular subjects—such as biology and clinical research—were predominant within it. To improve our representation of less frequent subjects such as social sciences and chemistry, we adopted a frequency weighting resampling technique Goloboff et al., 2003). This helped build a dataset that balanced the dominant with the minor subjects, using Formula ((2)

$$Weight_i = \frac{1}{1 + \lg freq_i} \tag{2}$$

Table 2
NB value range for author grouping.

Tier	1st Tier	2nd Tier	3rd Tier	4th Tier	Last Tier
NB values	[0,0.2)	[0.2,0.4)	[0.4,0.6)	[0.6,0.8)	[0.8,1]

Table 3
The disciplinary distribution of PLOS articles in our dataset.

Discipline	Frequency	Discipline	Frequency
Clinical Medicine	38,011	Physics	780
Biomedical Research	32,395	Social Sciences	689
Biology	9174	Chemistry	442
Psychology	2303	Professional Fields	313
Mathematics	2245	Unknown	88
Earth and Space	1114	Humanities	16
Engineering and Technology	795	Arts	1
Health	783		

where $freq_i$ is the number of publications in the subject, *i*. This formula reduces the sample number and weighting of subjects with more publications and increases the weight allocated to less widely published subjects.

We further stabilized our results by selecting only the top 12 disciplines for further investigation: clinical medicine, biomedical research, biology, psychology, mathematics, earth and space, engineering and technology, health, physics, social sciences, chemistry, and professional fields. Together, these subjects contributed 89,044 publications. To balance the sampling weight for each discipline, we found replacement papers, regenerating our dataset to comprise 70,000 publications in total. The subject distribution is shown in Table 4.

3.3.7. Null models used in this study

Null models are frequently adopted in network science studies (Blondel et al., 2008; Radicchi et al., 2008; Wang et al., 2013) to generate a dataset that can be compared with the experimental data to determine the presence of any random effects. In the social and medical sciences, this method is referred to as a randomized controlled trial (Fisher, 1936). In some scenarios, directly designing randomized controlled trials may be unachievable, so tools such as an (ER) random graph or configuration graph can be used to generate comparable network data (Lu et al., 2020). Thus, following the previous study, we used two null models to help validate our observations from the co-contributorship networks (see Fig. 4 in Lu et al., 2020).

4. Findings

In this section, we answer our four research questions. We first report the relationship between collaborators' byline orders and their contributions before comparing the byline order distributions of the three contributing roles in scientific collaborations. Finally, we detail how team size and disciplinary context influenced the byline order distributions of the three contributing roles.

4.1. Byline order generally reflects the contributions of collaborators

Figs. 3 and 4 show the contributions performed by the contributors according to their byline positions. Generally, collaborators with smaller normalized byline order values performed a higher variation and quantity of tasks, confirming the findings of other studies (Corrêa et al., 2017; Larivière et al., 2016). Specifically, collaborators from the first tier (i.e., those with a normalized byline order of [0,0.2)) undertook many more tasks than their colleagues on projects. They were particularly engaged in the first four common tasks: performing experiments, analyzing data, writing papers, and conceiving ideas. As first authors of the papers, they were less involved in revising and editing papers than other collaborators. This indicates that the authors from the first tier contributed the most to their scientific collaboration and therefore deserved the most credit for their efforts. However, collaborators in the final tiers sometimes performed more tasks such as contributing tools, editing drafts, and collecting data, confirming the reverse order of byline position in relation to contributions found by Corrêa et al. (2017). Nevertheless, while this reverse order pattern was detectable, the authors who were first in the byline usually contributed more, thus upholding the contribution-based author ordering rules.

4.2. Versatiles are more likely to be first authors, with other contributors following them in the byline

Fig. 5(A) shows the complementary cumulative distribution function (CCDF)[2] of the byline orders of the three types of collaborators compared against our null models prior to resampling the data. The graph demonstrates that versatiles were generally named first in

[2] https://en.wikipedia.org/wiki/Cumulative_distribution_function#Complementary_cumulative_distribution_function_(tail_distribution).

Table 4
Disciplinary distribution of the resampled dataset.

Discipline	Frequency	Ratio(%)	Discipline	Frequency	Ratio(%)
Professional Fields	7052	10.07	Earth and Space	5895	8.42
Chemistry	6548	9.35	Psychology	5838	8.34
Physics	6472	9.25	Mathematics	5749	8.21
Social Sciences	6471	9.24	Biology	4875	6.96
Health	6287	8.98	Biomedical Research	4293	6.13
Engineering and Technology	6241	8.92	Clinical Medicine	4279	6.11

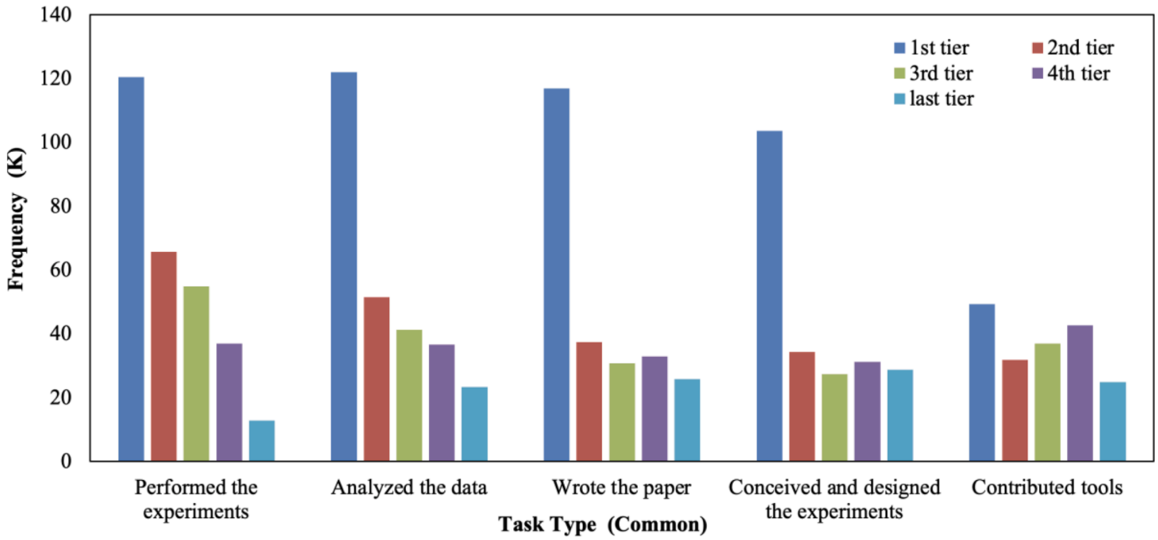


Fig. 3. Frequencies of common tasks performed by contributors in different tiers.

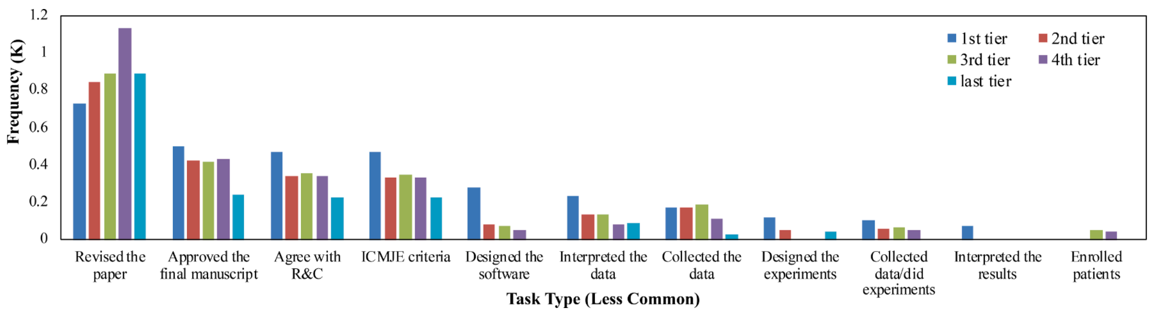


Fig. 4. Frequencies of less common tasks performed by contributors in different tiers.

bylines, ahead of teamplayers and specialists. This was confirmed by our null models, indicating that versatiles would receive larger normalized byline order values in randomized cases. Teamplayers usually followed their versatile colleagues in the middle of byline orders, with specialists' names at the end of the lists. These observations suggest that versatiles tend to lead studies or make a major contribution to their teams and are therefore usually assigned as first authors. Because specialists undertake various roles, including that of author, they were usually placed last in the bylines, indicating that their contributions were either marginal or involved coordinating research teams. The observation that versatiles—who usually make large contributions to the team—can also occasionally be last authors validates this interpretation.

To remove any potential biases caused by dominating subjects such as biology and medical sciences in our dataset, we resampled the data, increasing the weighting given to minor subjects. Fig. 5(B) shows that versatiles continued to assume a leading position among collaborators on the byline, confirming Lu et al.'s earlier findings (2020). Conversely, the names of specialists usually featured at the end of bylines, suggesting that their contributions tended to be marginal. Between versatiles and specialists on the byline lay those who usually perform more common tasks—the teamplayers. However, versatiles also occasionally appeared at the end of bylines, indicating their authoritative position as corresponding authors.

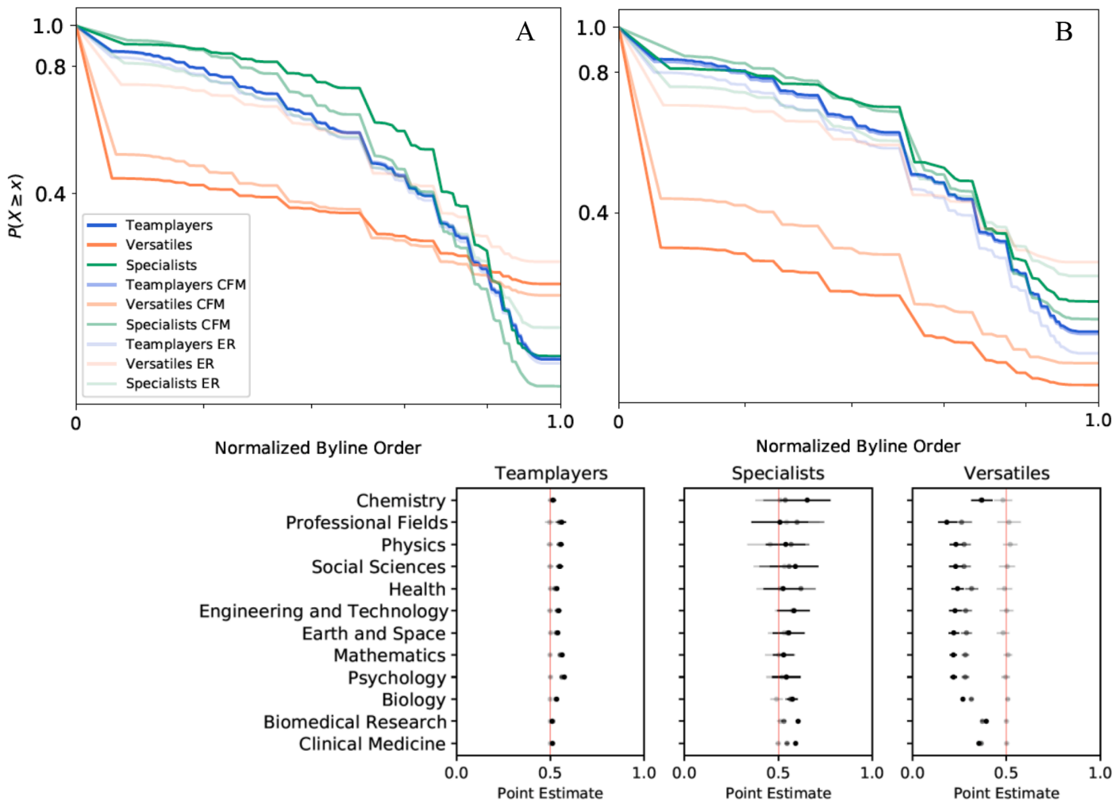


Fig. 5. (A) CCDF plot of byline orders by role against null models; (B) CCDF plot using resampled data; and (C) Point estimates of byline order at a 95% confidence interval of contributing roles by subject against null models at different degrees of transparency (CFM, 50%; ERM, 20%).

Similar findings were obtained from the point estimate analyses across disciplines shown in Fig. 5(C). Specifically, the names of most collaborators, i.e., teamplayers, usually occupied the center ground of the byline, with no obvious variation across subjects. Versatiles tended to place first in social sciences and physics bylines but were further behind in biomedical and clinical research, perhaps because relatively few versatiles are found in those domains. Larger variation in the byline order of specialists indicates their unique team functions as either leaders or assistants. Subjects with more specialist contributors were more likely to place these roles in different positions in their article bylines.

4.3. Versatiles can be both first and last authors in the byline while the other roles lie in the middle

Splitting our dataset by discipline, we plotted the byline order distributions of the collaborators against null models (see Fig. 6). Once more, the byline order distributions for each subject suggested that versatiles were most likely to be listed as first authors in the byline; teamplayers and specialists were both located behind the versatiles, and no significant differences in the distributions were detected. For instance, in social sciences, the versatiles' primary position was stronger than predicted by the null models, showing they were frequently listed among the leading positions in the bylines, such as either first or second author, whereas teamplayers tended to occupy the middle of these bylines and specialists appeared at the end of the bylines as last authors. A closer check suggested that the difference between teamplayers and specialists was marginal compared to the larger gap between versatiles and the other collaborators. Similar observations were identified in many other subjects, including psychology and mathematics. However, in a few disciplines, i.e., clinical medicine, biomedical research, and chemistry, versatiles could be listed as last authors. For instance, in clinical medicine, versatiles appeared behind teamplayers and specialists when the normalized byline order reached 0.8, indicating that they were more likely than the other two to be the last authors of the corresponding studies. This finding resembled those of biology and health, where versatiles were more likely to be placed at the end of the bylines than in other subjects such as social sciences.

4.4. Team size grouping signifies different forms of the division of labor and byline practice among subjects

To examine how team size affected the distributions, we further analyzed our dataset by the number of authors per article in the same discipline. We found that the distribution of contributing roles differed between subjects when team sizes increased. Two groups of subjects were identified. The first consisted of subjects where the three contributing roles were maintained in larger team sizes, including biology (as shown in Fig. 7), biomedical research (Fig. 8), clinical medicine, mathematics, earth and space, engineering &

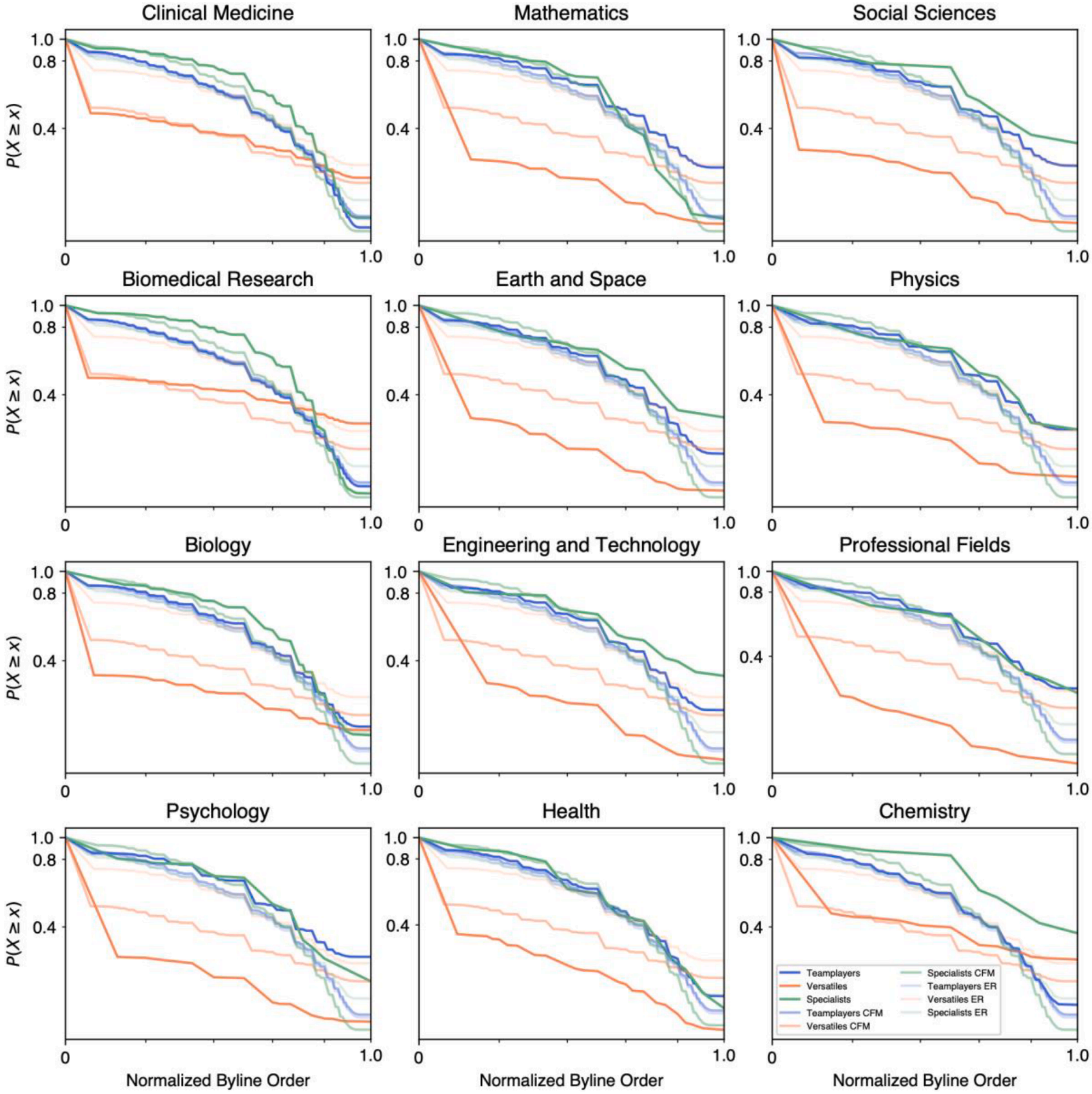


Fig. 6. CCDFs of normalized byline order by the subject against null models.

technology, and health.[3]

To exemplify this finding, in biology papers with no more than six contributors (common for most scientific collaborations; Lu et al., 2020), versatiles overwhelmingly tended to be first authors in the bylines, ahead of teamplayers and specialists. However, the plots indicated that the difference between them decreased in larger teams. Teams containing more than six members began to feature versatiles as well as the other two roles at the end of the bylines. The difference between the distribution of the three roles was lower in biology paper bylines (while not disappearing entirely) suggesting that the disciplinary teams might need a principal investigator to oversee the research process—and such people's names tend to be listed last in the byline.

The second group contains subjects in which some of the contributing roles begin to disappear with increasing team size. These subjects include chemistry, physics, professional fields, social sciences (see Fig. 9), and psychology (Fig. 10).

For example, as Fig. 9 illustrates, social sciences teams of six or fewer contained the three collaborator roles distributed across the byline, with versatiles as first authors, followed by teamplayers and specialists. However, as teams grew larger, some types of collaborators began disappearing from the byline: specialists first, then versatiles. Eventually, the collaboration contained only one type of contributor: teamplayers, meaning that no member completed project tasks independently from the others. In psychology, similar

[3] For conciseness, the findings are illustrated with a selection of the figures and plots. The data for all subjects will be made available on request.

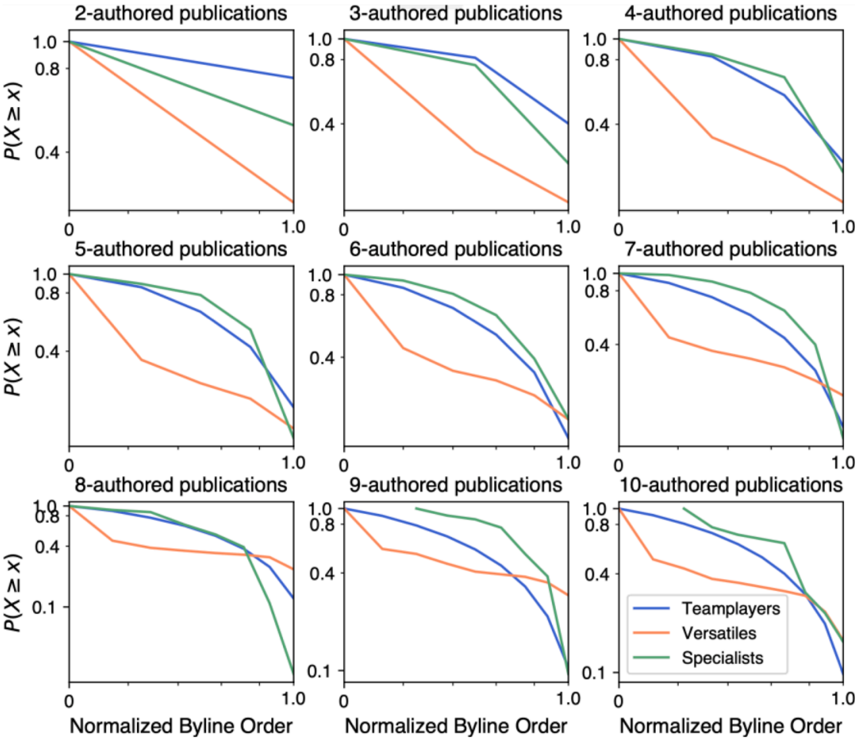


Fig. 7. Biology, a discipline where all contributor roles are retained as team size increases.

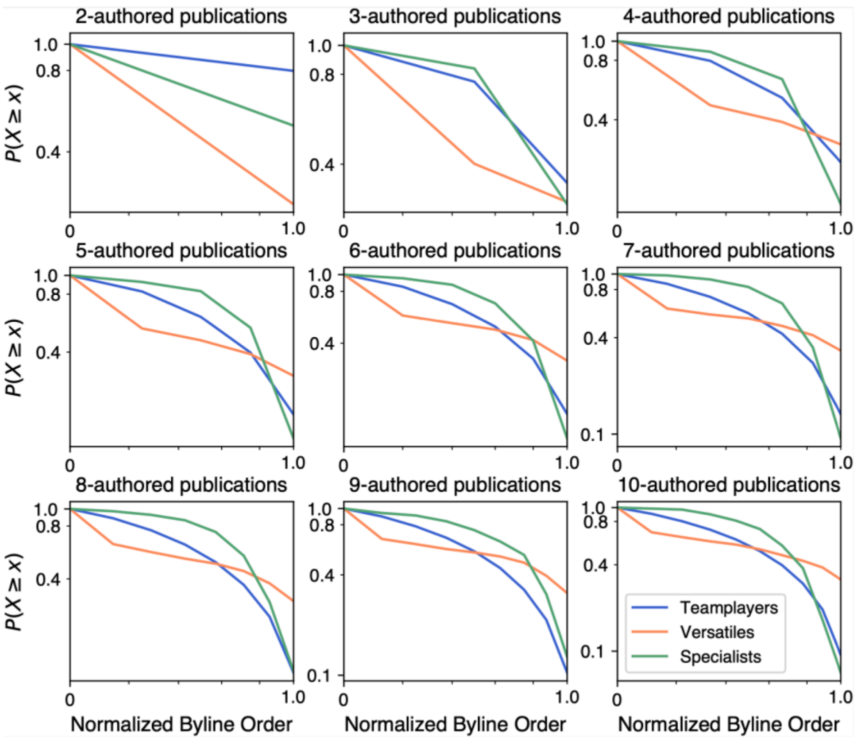


Fig. 8. Biomedical research, a discipline where all contributor roles are retained as team size increases.

patterns emerged as team size increased, with the specialist role vanishing from the teams, although unlike the previous example, the versatiles remained as last authors in the byline. Given the considerable contributions made by versatiles to their teams, they may be the corresponding authors whose names are found at the end of bylines.

We also observed that, in general, when teams grew larger, the boundaries between different contributor roles were vaguely reflected in the byline order. For instance, in biology teams of more than eight researchers, the distribution of versatiles resembled that of teamplayers and specialists. A similar tendency occurred in social sciences and psychology teams of seven or eight and was particularly evident in clinical medicine teams of more than four people, where teamplayers and specialists were most likely to be placed in the middle of the bylines while versatiles could be both first and last authors. Overall, we observed that larger team sizes were associated with much narrower margins between contributing roles.

5. Discussion and Conclusions

In an earlier study (Lu et al., 2020), we identified three types of contributorship, i.e., *versatiles, teamplayers*, and *specialists*. In this article, we have investigated how these are manifested in scientific collaboration by comparing the byline order distributions of the three types of contributorship across subjects and varying team sizes. For this, we generated a dataset comprising 105,192 articles with parsed author contribution statements and byline order information from the data we had previously collected (Lu et al., 2019b, 2020). By analyzing the author contribution statements, we identified the contributing roles of the authors using the co-contributorship network. We then examined the distribution of the three contributing roles (versatiles, teamplayers, and specialists) across subjects and team sizes. In what follows we discuss our key findings, the implications and limitations of the study, and concludes by highlighting potential directions for future research.

5.1. Key findings

First, we found that byline orders generally reflected the contributions made by collaborators, and thus represent contributorship. Versatiles, who are usually involved in multiple types of contribution and assume supervisory responsibilities, were more likely to be first authors but could also be the last name on the byline (Lu et al., 2020).

Second, we observed that larger team sizes exerted different effects on the three contributing roles, which were retained in subjects such as biology and biomedical research but were gradually subsumed into the teamplayer role in other subjects like social science and chemistry. It appears that the increase in team size foregrounds differences in disciplinary approaches to the division of labor and contributorship practices in scientific collaborations. For instance, biology research tasks appear to be more divisible while those of social science are less amenable to division. This suggests that collaborators in large biological research teams can be assigned more specific tasks, while their social science counterparts might be split into groups to perform tasks. Other possible interpretations might also center on the effects of an increasing number of part-time contracts in modern sciences (Milojević et al., 2018) or unethical authorship practices in some scientific collaborations (Bavdekar, 2012). However, these interpretations would require further investigation in future studies.

Finally, the similar distributions of the byline orders of the contributing roles in larger teams across disciplines indicated that the association between byline ordering and contributing roles weakens when team sizes grow. This may be explained by recent findings that byline ordering practices in larger research teams remain quite diverse—even within the same subject (Ali, 2021; Gómez-Ferri et al., 2019). It is also possible that the contribution of each researcher is diluted when their numbers increase, exerting a strong influence on the byline ordering process.

5.2. Implications

We will highlight two main implications that might concern the scientific community and policymakers. First, we observed that variable practices in authorship byline ordering might damage scientific communication more broadly, i.e., across domains and disciplines. One author's contribution to a collaborative study determines his or her recognition and ultimate status within the disciplinary community. However, the diversity of practices in byline ordering and author contribution disclosures that we uncovered indicates that evaluations are inconsistent across domains and disciplines, making it difficult for policymakers or outsiders to assess scientists' contributions to society, particularly when they have worked in larger teams. This may require authors to follow a more standardized authorship byline ordering practice and disclose their more detailed and domain-specific contributions. Second, our observations contribute to the ongoing debate concerning the pros and cons of huge teams. While scientific collaboration offers incontestable benefits to individuals, teams, and the science community, we observed that the division of labor in large teams is complex to coordinate (Becker & Murphy, 1992; Haeussler & Sauermann, 2020) and may disrupt the efficiency of scientific knowledge production. Moreover, larger teams increase the likelihood that some members might not be treated with due equity and care: teamplayers in particular might be placed randomly in the middle of a byline with contribution statements that are indistinguishable from those of their colleagues.

5.3. Limitations

Several limitations of the study must also be noted. First, it was difficult to obtain more accurate, long-term disciplinary classifications directly from PLOS publisher. This limited us to data from work published after 2015. Some temporal factors might not have

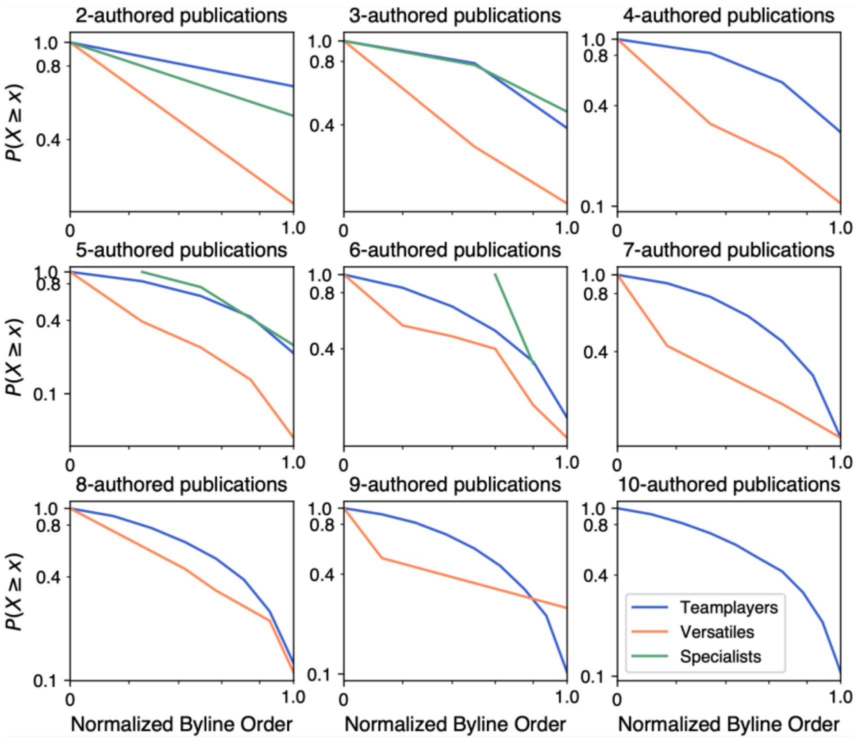


Fig. 9. Social sciences, a discipline where some contributor roles disappear as team size increases.

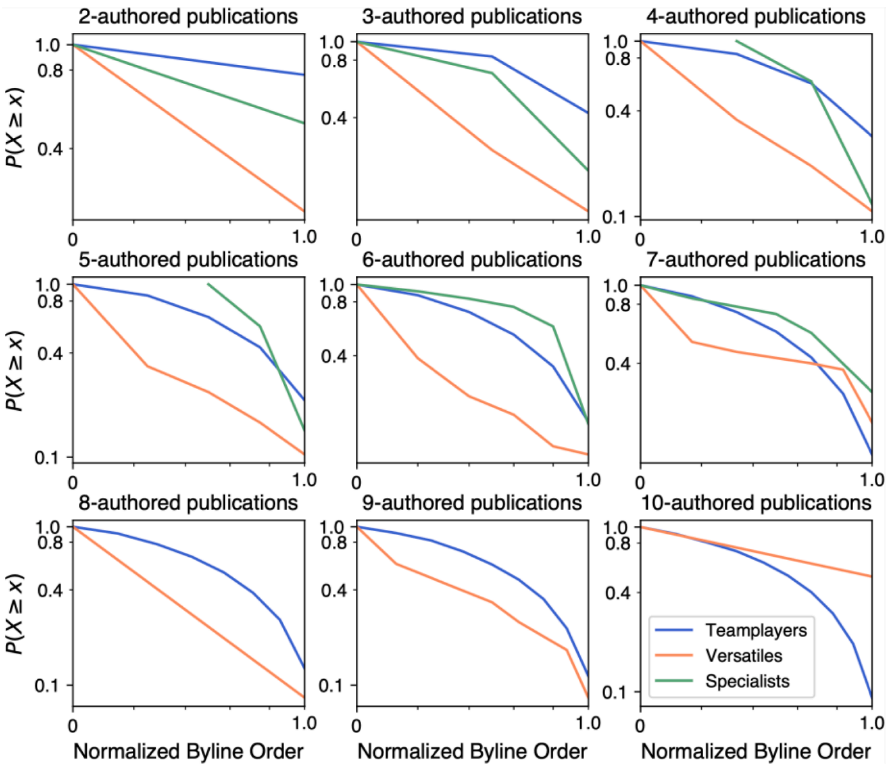


Fig. 10. Psychology, a discipline where some contributor roles disappear as team size increases.

been fully explained in this study. Second, PLOS specializes in the natural sciences rather than social sciences, humanities, and arts, which limits the generalizability of our findings to all disciplinary areas.

5.4. Future directions

Finally, we conclude our study by signifying some promising future directions for research. First, we would like to underline the potential value of co-contributorship as an instrument for measuring social capital more accurately in scientific collaboration from a social network perspective. For instance, when measuring structural social capital, Li et al. (2013) used three centrality metrics as the proxy. Examining contributorship in collaboration networks – particularly that of independent contributors – might allow structural social capital to be measured more accurately, enabling the relationship between the social capital of scholars and their career development to be investigated in greater depth while also advancing knowledge of scientific collaboration.

Second, future researchers could examine how the division of labor is organized in different disciplines, in order to promote the efficiency and efficacy of scientific collaboration. These modes of work vary in interesting ways: in the natural sciences, studies are characterized by step-by-step experimental designs which are well-suited to role specialization and tasks that can be carried out by individuals in relative independence. In contrast, the interdependent nature of social studies leaves fewer individualized tasks, which leads to another form of division of labor in which members work in small groups in larger teams.

Third, another area of investigation might examine how byline ordering practices impact the assessment of author contributions. So far, we have not yet arrived at a satisfactory way of assessing author contributions based on disclosure statements. An important first step would be for journals to require all submitted papers to include these statements. Moreover, given that author byline ordering practices vary by discipline (Bavdekar, 2012; Corrêa et al., 2017; Li & Yi, 2021), it will be useful to explore how byline ordering is linked to the assessment of author contributions in reality. These findings can further advance the disclosure of author contributions and lead to greater transparency when submitting scientific papers in the future.

Last, future studies could take a longitudinal perspective on contributing roles and the career development of those who fulfill them. At the individual level, it will be valuable to investigate whether and how collaborating scientists move from “survivor” to “thriver” in academia (Milojević et al., 2018). At the team level, given that larger teams tend to be less “disruptive” innovators (Wu et al., 2019), it could be promising to examine how the structure or the division of labor in teams evolves when the size and profile of teams increases.

CRediT authorship contribution statement

Chao Lu: Conceptualization, Methodology, Data curation, Formal analysis, Funding acquisition, Writing – original draft. **Chenwei Zhang:** Conceptualization, Methodology, Writing – review & editing. **Chengrui Xiao:** Methodology. **Ying Ding:** Supervision, Writing – review & editing.

Acknowledgement

This article is an outcome of the youth project "Study of Scientific Collaborators’ Scientific Effectiveness from The Perspective of Division of Labor" (ID: 72004054) supported by the National Natural Science Foundation of China and the project "The Causal Effect of Team Diversity on Team Performance" supported by the Fundamental Research Funds for the Central Universities (ID: B220201058). The authors appreciate the anonymous reviewers for their detailed comments for the improvement of this work.

References

Ali, M. J. (2021). No room for ambiguity: The concepts of appropriate and inappropriate authorship in scientific publications. *Indian Journal of Ophthalmology, 69*(1), 36–41.

Amjad, T., Ding, Y., Xu, J., Zhang, C., Daud, A., Tang, J., et al. (2017). Standing on the shoulders of giants. *Journal of Informetrics, 11*(1), 307–323.

Arroyo Moliner, L., Gallardo-Gallardo, E., & Gallo de Puelles, P. (2017). Understanding scientific communities: A social network approach to collaborations in talent management research. *Scientometrics, 113*(3), 1439–1462.

Bao, P., & Zhai, C. (2017). Dynamic credit allocation in scientific literature. *Scientometrics, 112*(1), 595–606.

Bavdekar, S. B. (2012). Authorship issues. *Lung India : Official Organ of Indian Chest Society, 29*(1), 76–80.

Becker, G. S., & Murphy, K. M. (1992). The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics, 107*(4), 1137–1160.

Bhandari, M., Guyatt, G. H., Kulkarni, A. V., Devereaux, P. J., Leece, P., Bajammal, S., et al. (2014). Perceptions of authors’ contributions are influenced by both byline order and designation of corresponding author. *Journal of Clinical Epidemiology, 67*(9), 1049–1054.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008.

Bu, Y., Ding, Y., Liang, X., & Murray, D. S. (2018). Understanding persistent scientific collaboration. *Journal of the Association for Information Science and Technology, 69*(3), 438–448.

Bu, Y., Wang, B., Chinchilla-Rodríguez, Z., Sugimoto, C. R., Huang, Y., & Huang, W. (2020). Considering author sequence in all-author co-citation analysis. *Information Processing & Management, 57*(6), Article 102300.

Burrows, S., & Moore, M. (2011). Trends in authorship order in biomedical research publications. *Journal of Electronic Resources in Medical Libraries, 8*(2), 155–168.

Chawla, D. S. (2019). Hyperauthorship: Global projects spark surge in thousand-author papers. *Nature*. https://www.nature.com/articles/d41586-019-03862-0.

Chen, C., Feng, J., Liu, X., & Yao, J. (2021). Leader humility, team job crafting and team creativity: The moderating role of leader–leader exchange. *Human Resource Management Journal, 31*(1), 326–340.

Chung, E., Kwon, N., & Lee, J. (2016). Understanding scientific collaboration in the research life cycle: Bio- and nanoscientists’ motivations, information-sharing and communication practices, and barriers to collaboration. *Journal of the Association for Information Science and Technology, 67*(8), 1836–1848.

Corrêa, E. A., Silva, F. N., da Costa, F. L., & Amancio, D. R. (2017). Patterns of authors contribution in scientific manuscripts. *Journal of Informetrics, 11*(2), 498–510.

Delfgaauw, J., Dur, R., & Souverijn, M. (2018). Team incentives, task assignment, and performance: A field experiment. *The Leadership Quarterly, 31*(3), Article 101241.

Feng, S., & Kirkley, A. (2020). Mixing patterns in interdisciplinary Co-authorship networks at multiple scales. *Scientific Reports, 10*, 7731.

Fisher, R. A. (1936). The design of experiments. *British Medical Journal, 1*(3923), 553–554.

Frische, S. (2012). It is time for full disclosure of author contributions. *Nature News, 489*(7417), 475.

Gazni, A., Sugimoto, C. R., & Didegah, F. (2012). Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the American Society for Information Science and Technology, 63*(2), 323–335.

Goloboff, P. A., Farris, J. S., Källersjö, M., Oxelman, B., Ramiacute; rez, M.n J., & Szumik, C. A. (2003). Improvements to resampling measures of group support. *Cladistics : The International Journal of the Willi Hennig Society, 19*(4), 324–332.

Gómez-Ferri, J., González-Alcaide, G., & LLopis-Goig, R. (2019). Measuring dissatisfaction with coauthorship: An empirical approach based on the researchers' perception. *Journal of Informetrics, 13*(4), Article 100980.

Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science, 308*(5722), 697–702 (New York, N.Y.).

Haeussler, C., & Sauermann, H. (2020). Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity. *Research Policy, 49*(6), Article 103987.

Kim, J., & Kim, J. (2015). Rethinking the comparison of coauthorship credit allocation schemes. *Journal of Informetrics, 9*(3), 667–673.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science, 46*(3), 417–435.

Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology, 66*(7), 1323–1332.

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.

Li, W., Aste, T., Caccioli, F., & Livan, G. (2019). Early coauthorship with top scientists predicts success in academic careers. *Nature Communications, 10*, 5170.

Li, W., & Yi, J. (2021). Alphabetical author order, intellectual collaboration and high-skilled migration. *The Economic Journal, 131*(635), 1250–1268.

Lu, C., Bu, Y., Dong, X., Wang, J., Ding, Y., Lariviere, V., et al. (2019a). Analyzing linguistic complexity and scientific impact. *Journal of Informetrics, 13*(3), 817–829.

Lu, C., Bu, Y., Wang, J., Ding, Y., Torvik, V., Schnaars, M., et al. (2019b). Examining scientific writing styles from the perspective of linguistic complexity. *Journal of the Association for Information Science and Technology, 70*(5), 462–475.

Lu, C., Zhang, Y., Ahn, Y., Ding, Y., Zhang, C., & Ma, D. (2020). Co-contributorship network and division of labor in individual scientific collaborations. *Journal of the Association for Information Science and Technology, 71*(10), 1162–1178.

Mattsson, P., Sundberg, C. J., & Laget, P. (2011). Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics, 87*(1), 99–105.

Milojević, S., Radicchi, F., & Walsh, J. P. (2018). Changing demographics of scientific careers: The rise of the temporary workforce. *Proceedings of the National Academy of Sciences, 115*(50), 12616–12623.

Ni, C., Smith, E., Yuan, H., Larivière, V., & Sugimoto, C. R. (2021). The gendered nature of authorship. *Science Advances, 7*, eabe4639.

Nielsen, M. W., & Andersen, J. P. (2021). Global citation inequality is on the rise. *Proceedings of the National Academy of Sciences, 118*(7), Article e2012208118.

Perneger, T. V., Poncet, A., Carpentier, M., Agoritsas, T., Combescure, C., & Gayet-Ageron, A. (2017). Thinker, soldier, scribe: Cross-sectional study of researchers' roles and author order in the annals of internal medicine. *BMJ Open, 7*(6), Article e013898.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences, 105*(45), 17268–17272.

Rahman, M. T., Regenstein, J. M., Kassim, N. L. A., & Haque, N. (2017). The need to quantify authors' relative intellectual contributions in a multi-author paper. *Journal of Informetrics, 11*(1), 275–281.

Sauermann, H., & Haeussler, C. (2017). Authorship and contribution disclosures. *Science Advances, 3*(11), Article e1700404.

Teixeira da Silva, J. A. (2021). Multiple co-first authors, co-corresponding authors and co-supervisors: A synthesis of shared authorship credit. *Online Information Review, 45*(6), 1116–1130.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science, 342*(6157), 468–472 (New York, N.Y.).

Wang, D., & Barabási, A.-. L. (2021). *The science of science*. Cambridge University Press.

Wang, D., Song, C., & Barabási, A.-. L. (2013). Quantifying long-term scientific impact. *Science, 342*(6154), 127–132 (New York, N.Y.).

Wang, W., Ren, J., Alrashoud, M., Xia, F., Mao, M., & Tolba, A. (2020). Early-stage reciprocity in sustainable scientific collaboration. *Journal of Informetrics, 14*(3), Article 101041.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378–382.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039 (New York, N.Y.).

Xu, J., Ding, Y., Song, M., & Chambers, T. (2016). Author credit-assignment schemas: A comparison and analysis. *Journal of the Association for Information Science and Technology, 67*(8), 1973–1989.

Xu, S., Li, L., Hao, L., An, X., & Yang, G. (2021). An author interest discovery model armed with authorship credit allocation scheme. In K. Toeppe, H. Yan, & S. K. W. Chu (Eds.), *Diversity, divergence, dialogue* (pp. 199–207). Springer International Publishing.

Yang, S., Wolfram, D., & Wang, F. (2017). The relationship between the author byline and contribution lists: A comparison of three general medical journals. *Scientometrics, 110*(3), 1273–1296.

Zhang, M., Zhang, G., Liu, Y., Zhai, X., & Han, X. (2020). Scientists' genders and international academic collaboration: An empirical study of Chinese universities and research institutes. *Journal of Informetrics, 14*(4), Article 101068.

Zhao, Z., Bu, Y., & Li, J. (2021). Characterizing scientists leaving science before their time: Evidence from mathematics. *Information Processing & Management, 58*(5), Article 102661.