Relationship Between Paper Authorship Roles and Novelty from a Gender Perspective: Evidence from 81,137 PLOS ONE Articles*

Jiaqi Zeng, Yi Zhao and Chengzhi Zhang*

Department of Information Management, Nanjing University of Science and Technology 210094 Nanjing, China

Abstract

The division of labor within a research paper plays a crucial role in fostering efficient collaboration and knowledge innovation. The authors' engagement in different contributions influences the integration of specialized knowledge, the formation of diverse perspectives, and the stimulation of creativity, which in turn impact the novelty level of the paper. However, previous studies have lacked depth in exploring the relationship between paper division of labor and novelty, and have overlooked potential gender differences. This study, based on 81,137 papers from PLOS ONE, investigates the correlation between authors' contributions engagement, contributions engagement of authors of different genders, and paper novelty. The results show that, in the Writing-original draft preparation, Writing-review & editing, Methodology, and Software, a higher proportion of author participation is associated with a higher likelihood of the paper achieving greater novelty. In terms of gender differences, women are more likely to participate in the Investigation, Data curation, Formal analysis, and Writing-original draft preparation, while men tend to play a more prominent role in Supervision, Resources, Funding acquisition, Conceptualization, and Software. Furthermore, the study shows that, regardless of gender, a greater proportion of participation in the Writing-original draft preparation, Writing-review & editing, and Software is significantly associated with higher paper novelty. However, only for male authors, a greater proportion of participation in Methodology, Visualization, and Funding acquisition is associated with higher paper novelty.

Keywords

Author division of labor, Novelty, Gender differences

1. Introduction

Scientific collaboration is defined as "a concerted effort by researchers to achieve a common goal of generating new scientific knowledge" [1]. It is widely recognized that scientific collaboration has a positive impact on academic success [2]. Scientific collaboration relies heavily on the division of labor, which effectively integrates scholars' unique expertise, skills, and research experience, fostering innovative thinking. Understanding how to achieve high-quality scientific collaboration through division of labor is of great significance [3]. Novelty assessment is a crucial aspect of academic quality evaluation. Previous studies have evaluated paper novelty from multiple perspectives, using methods such as citation analysis, entity analysis, and semantic analysis [4, 5, 6]. However, these studies primarily focus on novelty assessment and lack in-depth exploration of the underlying factors influencing paper novelty. Scientific division of labor and collaboration, by bringing together diverse perspectives and expertise, provide the potential for generating high-novelty research outcomes [7, 8]. Contribution engagement, as a crucial indicator of the degree of author involvement and effort in each contribution, has not been adequately explored in terms of its impact on paper novelty. Therefore, this paper aims to empirically investigate the potential relationship between paper division of labor and paper novelty. This study is of great significance for scientific team formation, promoting a reasonable division of labor in papers, and driving scientific innovation. We aim to answer the following research question:

* Joint Workshop of the 2th Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (IMSC2024), Dec 20th, 2024, Hong Kong, China and Online

*Corresponding author.

✉zengjq@njust.edu.cn (J. Zeng); yizhao93@njust.edu.cn (Y. Zhao); zhangcz@njust.edu.cn (C. Zhang)

0009-0005-4161-3799 (J. Zeng); 0000-0003-1050-7051 (Y. Zhao); 0000-0001-9522-2914 (C. Zhang)

RQ1: Is author's participation in different research contributions in papers correlated with paper novelty?

Previous research on author division of labor based on author contribution statements has revealed significant differences in division of labor among scholars of different genders. Haeussler et al., based on data from 12,964 papers from PLOS ONE, found that women are more likely to participate in experimental rather than conceptual activities [7]. Research on gender disparities in novelty, such as that conducted by Liu et al., has found that biomedical doctoral dissertations written by women exhibit lower scientific novelty compared to those written by men [9]. While previous studies have documented significant differences in division of labor between male and female scholars, and have observed differences in the novelty of their publications, none have delved into the question of whether gender disparities in division of labor influence the novelty of research papers. Understanding whether the division contribution engagement of authors of different genders is related to paper novelty is essential for investigating gender differences in academia and uncovering gender-related factors in the creation of high-novelty papers. Therefore, this paper further explores the following research question:

RQ2: Is the research contribution engagement of authors of different genders correlated with paper novelty?

2. Methodology

PLOS ONE is a multidisciplinary open access journal that supports the development of knowledge dissemination. Therefore, this study selects articles published in PLOS ONE as the research data, retrieving a total of 124,688 papers published between 2016 and 2024 from the PLOS ONE journal website[1]. Focusing on gender differences in the division within collaborations, we exclude 1,860 single-authored papers. Author gender identification is conducted using the Genderize.io tool[2], and papers with incomplete gender identification are removed (11,985 papers), leaving 110,843 papers for analysis. Lin et al. developed the SciSciNet dataset, which encompasses over 134 million scientific publications and millions of external links related to funding and public uses, providing metrics such as paper novelty [10]. We use the paper DOIs to match PLOS ONE papers with SciSciNet records, obtaining the tail novelty metric (Atyp_10pct_Z), developed by Uzzi et al. [4]. This metric measures novelty based on the commonality of co-cited journal pairs in the references. A lower Atyp_10pct_Z value indicates higher paper novelty. We successfully match 81,137 papers from 2016 to 2021.

Since 2016, PLOS ONE has adopted the CRediT contribution taxonomy, encompassing 14 research contributions [3]. Based on 81,137 PLOS ONE papers, this study investigates the relationship between the author participation rate in each contribution (as shown in Formula 1), the male author participation rate (as shown in Formula 2), the female author participation rate (as shown in Formula 3), in relation to the novelty of the paper (Atyp_10pct_Z). The regression model controls for team size (Teamsize), the proportion of contribution categories involved in each paper (All_Contribution_per, as shown in Formula 4), the publication year (Fixed year), the proportion of female/male authors in the paper (Per_f/Per_m).

$$P\_i = \frac{Contribution\_i\_authors}{Total\_authors} \tag{1}$$

$$P\_m\_i = \frac{Contribution\_i\_male\_authors}{Total\_authors} \tag{2}$$

$$P\_f\_i = \frac{Contribution\_i\_female\_authors}{Total\_authors} \tag{3}$$

$$All\_Contribution\_per = \frac{CRediT\_contributions}{14} \tag{4}$$

Where *i* refers to one of the 14 contributions defined in the CRediT taxonomy: 'Conceptualization', 'Data curation', 'Formal analysis', 'Funding acquisition', 'Investigation', 'Methodology', 'Project Administration', 'Resources', 'Software', 'Supervision', 'Validation', 'Visualization', 'Writing-original

[1] https://journals.plos.org/plosone/
[2] https://genderize.io/
[3] https://credit.niso.org/

draft preparation', and 'Writing-review & editing'. *Contribution_i_authors* represents the number of authors contributing to *i* per paper. *Total_authors* represents the total number of authors per paper. *Contribution_i_male_authors and Contribution_i_female_authors* signify the respective counts of male and female authors contributing to *i* per paper, and *CRediT_contributions* represent the contribution categories of CRediT involved in each paper.

3. Results

This study investigates whether gender-based preferences exist in the division of labor among authors of PLOS ONE publications. It explores the relationship between author contribution engagement and paper novelty, specifically examining whether this relationship differs between male and female authors.

3.1. Gender Differences in Author Contribution Engagement

The 81,137 papers from PLOS ONE involve a total of 534,898 authors, with 208,733 female authors (39%) and 326,165 male authors (61%). Figure 1 presents the author participation rates in contributions and the difference in participation rates between genders. In Figure 1, All represents the proportion of authors participating in *i* to the total number of authors in all papers. F-M represents the difference in participation rate between male and female authors. It is calculated as the proportion of female participation in *i* (i.e., the total number of female authors performing *i* in all papers / the total number of female authors) minus the proportion of male participation in *i* (i.e., the total number of male authors performing *i* in all papers / the total number of male authors). Here, *i* belongs to the 14 contributions of CRediT.

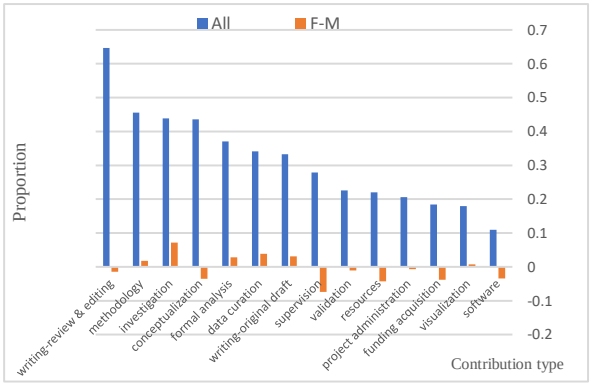

Figure 1: Author Participation Rates in research contributions and the Difference in Participation Rates Between Genders

Figure 1 reveals that Writing-review & editing, Methodology, and Investigation are the contributions with the highest participation rates, while Software, Visualization, and Funding acquisition have the lowest participation rates. The F-M difference reveals that women are more likely than men to participate in Investigation, Data curation, Formal analysis, and Writing-original draft preparation in collaborative research in PLOS ONE papers. Conversely, men tend to participate more in Supervision, Resources, Funding acquisition, Conceptualization, and Software. These findings suggest the presence of gender biases in the division of labor in research publications. This study employs regression analysis to examine the relationship between author participation rates in different contribution roles and paper novelty, addressing RQ1. The dependent variable, Atyp_10pct_Z, represents paper novelty, with lower values indicating higher novelty. A multiple linear regression model is used to explore the association between author division of labor engagement and paper novelty. The regression results are presented in Model 1 of Table 1. The variable descriptions are provided in the Methodology section. The results indicate that papers with a higher proportion of authors participating in the Writing-original draft preparation, Writing-review & editing, Methodology, and Software contributions are more likely to exhibit higher novelty. Conversely, papers with a higher proportion of authors participating in

Resources, Data curation, Investigation, Validation, and Supervision are associated with lower novelty.

3.2. The Relationship Between Author Contribution Engagement and Paper Novelty: A Gendered Analysis

To address *RQ2*, two additional multiple linear regression analyses (Models 2 & 3 in Table 1) are conducted, focusing specifically on the relationship between the participation rates of female and male authors in various contribution roles and paper novelty. The regression results, presented in Table 1, reveal that a greater proportion of female author participation in Writing-original draft preparation, Writing-review & editing, and Software contributions is significantly associated with higher paper novelty. Conversely, a greater proportion of female author participation in Resources, Data curation, Formal analysis, and Investigation is associated with lower paper novelty. For male authors, a greater proportion of participation in Visualization, Writing-original draft preparation, Writing-review & editing, Methodology, Software, and Funding acquisition is significantly associated with higher paper novelty. However, a greater proportion of male author participation in Resources, Data curation, Investigation, Validation, and Supervision is associated with lower paper novelty.

The findings of this section not only confirm the correlation between author participation rates in different contribution roles and paper novelty, but also reveal the differences in the relationship between the participation rates of authors of different genders and paper novelty. Comparing the relationship between male and female author participation rates and paper novelty, it is found that only for males, participation in Visualization, Methodology, and Funding acquisition contributions is positively correlated with paper novelty. This finding provides a new research perspective for exploring the potential relationship between author gender and innovative outputs in scientific collaboration.

Table 1
Regression Analysis of Author Contribution Participation and Paper Novelty

Model	(Model 1)	Model	(Model 2)	Model	(Model 3)
VARIABLES	Atyp_10pct_Z	VARIABLES	Atyp_10pct_Z	VARIABLES	Atyp_10pct_Z
Per_f	1.128***	Per_f	0.670	Per_m	0.876
	(0.437)		(0.861)		(0.669)
Teamsize	-0.104***	Teamsize	-0.00411	Teamsize	-0.101***
	(0.0344)		(0.0308)		(0.0320)
All_Contribution_per	-4.209***	All_Contribution_per	-5.179***	All_Contribution_per	-3.304***
	(0.919)		(0.718)		(0.800)
P _conceptualization	-0.477	P_f_conceptualization	0.403	P_m_conceptualization	-0.995
	(0.468)		(0.844)		(0.631)
P_resources	1.651***	P_f_resources	2.868***	P_m_resources	1.419**
	(0.500)		(0.933)		(0.666)
P_visualization	-0.688	P_f_visualization	0.684	P_m_visualization	-1.734**
	(0.531)		(0.941)		(0.737)
P_ writing-original draft preparation	-1.767***	P_f_writing-original draft preparation	-2.532***	P_m_ writing-original draft preparation	-1.665***
	(0.477)		(0.855)		(0.633)
P_writing-review & editing	-1.547***	P_f_writing-review & editing	-1.513**	P_m_writing-review & editing	-2.446***
	(0.381)		(0.727)		(0.541)
P_data curation	2.932***	P_f_data curation	2.937***	P_m_data curation	4.072***
	(0.426)		(0.751)		(0.600)
P_formal analysis	0.370	P_f_formal analysis	1.786**	P_m_formal analysis	0.0255
	(0.470)		(0.829)		(0.641)
P_investigation	0.917**	P_f_finvestigation	1.691**	P_m_investigation	0.936*
	(0.389)		(0.710)		(0.551)
P_methodology	-1.140***	P_f_methodology	-0.807	P_m_methodology	-1.664***
	(0.426)		(0.768)		(0.593)
P_software	-5.201***	P_f_software	-5.365***	P_m_software	-6.243***
	(0.603)		(1.154)		(0.777)
P_validation	1.041**	P_f_validation	0.785	P_m_validation	1.516**
	(0.459)		(0.848)		(0.636)
P_funding acquisition	-0.402	P_f_funding acquisition	1.238	P_m_funding acquisition	-1.322*
	(0.567)		(1.022)		(0.729)
P_project administration	-0.349	P_f_project administration	-0.863	P_m_project administration	-0.273
	(0.584)		(1.004)		(0.767)
P_supervision	1.215**	P_f_supervision	-0.298	P_m_supervision	2.143***
	(0.543)		(0.978)		(0.683)
Fixed year	Yes	Fixed year	Yes	Fixed year	Yes
Constant	2.309***	Constant	1.141	Constant	1.872**
	(0.771)		(0.720)		(0.780)
Observations	81,137	Observations	81,137	Observations	81,137
R-squared	0.004	R-squared	0.003	R-squared	0.004

Note: *** p<0.01, ** p<0.05, * p<0.1 (***, **, * indicate significant at 1%, 5%, and 10% significance levels, respectively.

4. Conclusion and future works

This study investigates the relationship between gender differences in author division of labor and paper novelty. Our findings indicate that women are more likely than men to participate in contribution roles related to analysis, data curation, and writing. Men, on the other hand, are more likely to participate in contribution roles involving Supervision, Funding acquisition, and Conceptualization of the paper. The results show that papers with a higher proportion of authors participating in the Writing-original draft preparation, Writing-review & editing, Methodology,

and Software contributions are more likely to exhibit higher novelty. A greater proportion of both male and female authors' participation in Writing-original draft preparation, Writing-review & editing, and Software contributions is significantly associated with higher paper novelty. Additionally, for male authors, a greater proportion of participation in Visualization, Methodology, and Funding acquisition is significantly associated with higher paper novelty. The findings of this study provide reference suggestions for optimizing the division of labor in research teams to achieve the production of highly novel papers.

This study's data is confined to PLOS ONE publications. Future research could extend this analysis to other academic journals, investigating the relationship between author contribution engagement and paper novelty across different publication venues. Additionally, this study employs the Atyp_10pct_Z metric, developed by Uzzi et al. [4], to assess paper novelty. Future research could explore this relationship using alternative novelty metrics. Furthermore, while this study identifies a correlation between author contribution engagement and paper novelty, and observes differences between genders, future research could delve into the causal relationships underlying this connection between division of labor engagement and paper novelty.

Acknowledgements

The paper is presented at the second Workshop on "Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data" at 2024 ACM/IEEE Joint Conference on Digital Libraries (JCDL). This work was partially supported by the National Natural Science Foundation of China (No. 72074113).

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4 in order to correct grammatical errors, typos, and other writing mistakes. After using this tool, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] J.S. Katz, B.R. Martin, What is research collaboration? , Research Policy, 26 (1997) 1-18.
[2] H. Shen, J. Xie, W. Ao, Y. Cheng, The continuity and citation impact of scientific collaboration with different gender composition, Journal of Informetrics, 16 (2022) 101248.
[3] C. Lu, Y. Zhang, Y.Y. Ahn, Y. Ding, C. Zhang, D. Ma, Co-contributorship network and division of labor in individual scientific collaborations, Journal of the Association for Information Science and Technology, 71 (2020) 1162-1178.
[4] B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact, Science, 342 (2013) 468-472.
[5] S. Shibayama, D. Yin, K. Matsumoto, Measuring novelty in science with word embedding, PloS one, 16 (2021) e0254034.
[6] Z. Luo, W. Lu, J. He, Y. Wang, Combination of research questions and methods: A new measurement of scientific novelty, Journal of Informetrics, 16 (2022) 101282.
[7] C. Haeussler, H. Sauermann, Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity, Research Policy, 49 (2020) 103987.
[8] M. Franceschet, A. Costantini, The effect of scholar collaboration on impact and quality of academic papers, Journal of informetrics, 4 (2010) 540-553.
[9] M. Liu, Z. Xie, A. J. Yang, C. Yu, J. Xu, Y. Ding, Y. Bu, The prominent and heterogeneous gender disparities in scientific novelty: Evidence from biomedical doctoral theses, Information Processing & Management, 61 (2024) 103743.
[10] Z. Lin, Y. Yin, L. Liu, D. Wang, SciSciNet: A large-scale open data lake for the science of science research, Scientific Data, 10 (2023) 315.