TYPE：Research in Progress

Authorship, inventorship and division of labor in innovative research: an analysis of paper-patent pairs

Philippe Mongeon

philippe.mongeon@umontreal.ca
École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal (Canada)
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Montréal (Canada)

Abstract

Collaboration has long become the norm in science, and research is performed by increasingly large teams. Collaboration can take diverse forms, bring together researchers from different fields, status, and skills, and may lead to different patterns of tasks division between team members. This complicates the attribution of credit to individual authors, undermines research evaluation processes, and potentially reduces the efficacy of the reward system of science. This work in progress investigates division of labor and among team members using discoveries that were disclosed in both a patent and a paper, both document thus forming a paper-patent pair (PPP). Since the inventor status on a patent is attributed, in principle, for specific types of contribution (namely, those of a conceptual nature), comparing authors and inventors lists of PPPs can provide new insights on the division of labor and credit attribution practices in scientific research. This work analyzes the relationship between the author and inventor lists of 1,568 PPPs from five disciplines. We find that the individuals who are both authors and inventors are polarized towards the first and last author position. However, there appears to be disciplinary differences in the distribution of inventors in the various positions of the authors list.

Conference Topic

Collaboration and division of labor
Authorship
Inventorship
Paper-patent pairs

Introduction

Science is increasingly collaborative (Wuchty, Jones, & Uzzi, 2007), which leads to an increase in the number of authors of scientific papers. The longer author list, but the diverse types of collaboration (Maienschein, 1993), team compositions (Smith, 1971), and ways tasks are divided among team members (Larivière et al., 2016) makes the attribution of credit and responsibility to individuals an increasingly complex task (Rennie, Yank, & Emanuel, 1997). This is especially true for authors who are in the middle of the list (Bennett & Taylor, 2003; Shapiro, Wenger, & Shapiro, 1994). In this context, it is very difficult to deduct the type and extent of individual authors' contribution based on their position in the byline. This is problematic for researchers whose career advancement largely depends on the credit they obtain for their work (Birnholtz, 2006; Cronin, 1996). It also undermines the efficiency of the reward system of science which requires that the best researchers be properly identified and rewarded (Merton, 1957). The importance of these issues has led to many studies that investigated authorship practices, or more specifically the relationship between individual contribution and authorship using different datasets and using both quantitative and qualitative methods.

Research in some disciplines may sometimes lead to technological innovations for which the researchers may choose to file a patent application. Patents are sometimes used as indicators of the usefulness or the inventiveness of research (Meyer, 2003), and because patenting is often highly valued in research evaluations, inventorship becomes, like authorship, a way for researchers to build a reputation and advance their career (Desrochers et al., in press). The resulting paper-patent pair (PPP) can be a useful tool to analyze the type and substantiality of the tasks performed by the individual authors, and on credit attribution practices within research groups. The value of PPPs for investigating authorship practices lies in the fact that publishing and patenting operate in different realms and come with their distinct sets of rules and norms (Packer & Webster, 1996). For example, not all publishable research findings are patentable, and vice-versa. Similarly, contributions typically rewarded with authorship (e.g., technical contributions, data collection, drafting of the manuscript) are often not sufficient to be named inventor on a patent (Haeussler & Sauermann, 2013). Indeed, to be inventor, one must have conceived the invention (In re Hardee, 223 USPQ 1122, 1123; Comm'r Pat. 1984). Also, formulating desired outcomes is not sufficient: the inventor must have developed a non-obvious and concrete solution (Ex parte Smernoff, 215 USPQ 545, 547; Bd. App. 1982). In this context, comparing the inventor list and the author list of a patent and a paper that report the same discovery may indicate which of the authors have performed the conceptual work, and which have performed the technical work or other types of task that do not, in principle, lead to inventorship.

A few previous studies analyzed the difference between the inventors and authors list of PPPs. Looking at a set of 40 PPPs, Ducor (2000) found that the first and last authors were more likely than middle authors to be named inventors. He also found that last authors were more often listed as inventors than first authors. Using survey data from 2000 life scientists who self-reported having both published and patented a research finding, Haeussler and Sauermann (2013) found that contributions that are conceptual in nature are strong predictors or both authorship and inventorship, but that other types of contributions lead to authorship but not necessarily to inventorship. They also found that past scientific accomplishments is a predictor of authorship and that hierarchical position is a predictor of inventorship, but not authorship. Our study complements the previous research by identifying PPPs using worldwide paper and patent data covering all discipline for a large-scale comparison of authorship and inventorship practices. We analyze the relationship individual researchers' position in the author list and their inclusion in the inventors list of the associated patent. More specifically, the present paper provides answers to the following research questions:

- RQ1. What is the difference between the average number of authors and number of inventors?
- RQ2. What is the relationship between the position in the authors' list and the probability of also being listed as an inventor on the patent?

While these two questions are the focus of the preliminary results presented in this paper, they will be complemented by the following third research question in further steps of the study:

- RQ3. What is the relationship between the authors' characteristics, such as academic age, gender, productivity and scientific impact on the probability of also being listed as an inventor on the patent?

Data and methods

We collected all patent applications filed between 2001 and 2013 at the United States Patents and Trademark Office, and all articles published between 2000 and 2015 and indexed in Clarivate Analytics' Web of Science (WoS). The articles published before 2000 are not considered since under the United States patent law, inventions published more than 12 months prior to the patent application are considered to be part of the public domain and are therefore not patentable. Also, only articles for which the abstract is available were included, as the abstract is used for the pairing of the papers and patents.

To form PPPs, we first paired all articles and patents with at least one author-inventor match, based on the last name and first initial, which yielded more than 20 million potential pairs. Then, similarly to previous work by Magerman, Van Looy and Debackere (2015), we measured the content similarity between the articles (title and abstract) and patents (title and description). The difference between our method and that of Magerman et al. (2015) is that we used noun phrases instead of words to calculate the similarity between each PPPs. For both documents forming a potential pair, we divide the number of matching noun phrases by the total number of noun phrases in the title and abstract. The average of these two values is the similarity score of the pair.

To validate the matched pair, we manually verified all cases where the first name of the author was available in the WoS record. A match is considered a true positive when the first name of the author matches the first name of the inventor. The results of this validation are presented in Table 1.

Table 1. Validation of the author-inventor matches for cases where the full first names of the authors are available

Similarity range	*Number of author-inventors matches*	*True positives*	*False positives*	*Error rate*
≥ .90	31	31	0	0.0%
.80-.89	177	176	1	0.6%
.70-.79	176	174	2	1.1%
.60-.69	821	770	0	0.0%
.50-.59	1,116	1,085	0	0.0%
.40-.49	1,873	1,854	19	1.0%
.30-.39	3,647	2,758	889	24.4%
.20-.29	2,360	1,421	939	39.8%

We see that very few false positives occurred for PPPs with a similarity score of 0.40 or above (less than 1% overall), while 24.4% and 39.8% of author-inventor matches were false positives at a 0.30-0.39 and 0.20-0.30 similarity ranges, respectively. Therefore, we chose to limit the following analysis to all PPPs with a similarity of 0.40 or above, including those for which the full names of the authors were not available. We manually verified each PPP to ensure that no false negative remained in the author-inventor matching, and then excluded from the analysis all PPPs for which at least one inventor is not in the authors list to ensure that all contributors are included in the analysis. When multiple patent matched with a single paper, two criteria were used to select a single best matching patent: 1) we kept the pair with the highest number of authors who are also inventors on the patent, and 2) we keep the PPPs with the highest similarity score. We also limited the analysis to articles with 2 or more authors since the goal is to analyze division of labor and credit attribution among team members. One outlier, a paper with 161 authors, was also excluded from the analysis. The final dataset comprises 1,568 PPPs from five disciplines (see Table 2 in the result section).

Results

Table 2 presents the number of PPPs found for each discipline, the average number of authors, the average number of inventors, and the average inventors/authors ratio. Note that a small number PPPs were found in some of the fields not figuring in table 2 (e.g., Biology, Professional

Fields and Earth and Space). We excluded these PPPs from the analysis because their number was too small. Engineering and Technology, perhaps unsurprisingly, is the discipline with the most PPPs (N= 522) while the number of PPPs found for other disciplines ranges between 229 in Clinical Medicine to 279 in Biomedical Research. The results show that the average number of authors varies more between disciplines than the average number of inventors. Thus, the inventors/authors ratio ranges from 47.2% in Clinical Medicine, to 81.3% in Engineering and Technology, which suggests that important disciplinary differences exists in terms of authorship practices and division of labor among members of the research teams.

Table 2. Average number of authors, number of inventors, and inventors/authors ratio by discipline

Discipline	*Number of PPPs*	*Avg. number of authors*	*Avg. number of inventors*	*Avg. inventors/authors ratio*
Biomedical Research	279	5.96	2.62	55.7%
Chemistry	267	4.53	2.39	62.8%
Clinical Medicine	229	6.06	2.24	47.2%
Engineering and Technology	522	3.35	2.56	81.3%
Physics	271	3.92	2.62	74.3%
Total	1,568	4.51	2.51	67.4%

We now investigate the relationship between an individual's position in the author list and the likelihood also being listed as an inventor on the patent. If we divide the authors into three groups (first authors, middle authors, and last authors), we see that the inventors/authors ratio is smaller for middle authors than for first and last authors in all disciplines (Figure 1). We also observe that neither the first, or last authors are systematically listed as inventors on the matching patents. Moreover, last authors are most likely to be inventors than first authors in Biomedical Research, Clinical Medicine and Chemistry (although the difference seems negligible in the case of Chemistry), while in the first authors are more likely than last authors to be inventors in Physics and in Engineering and Technology.

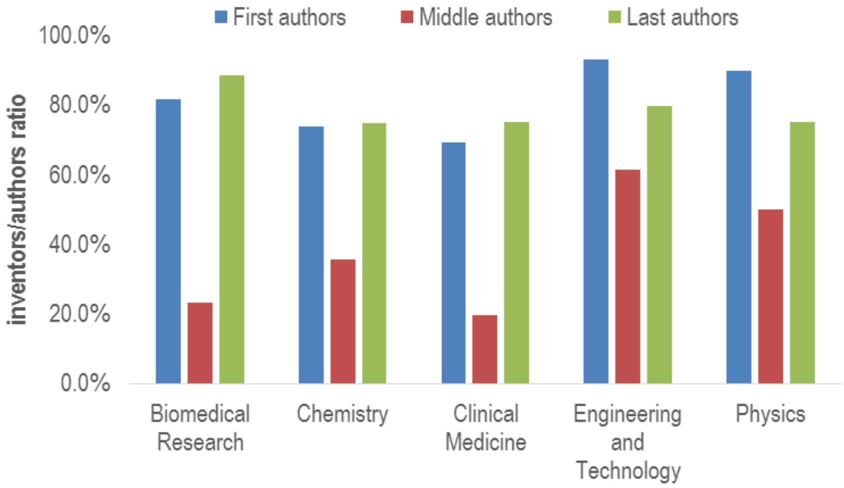
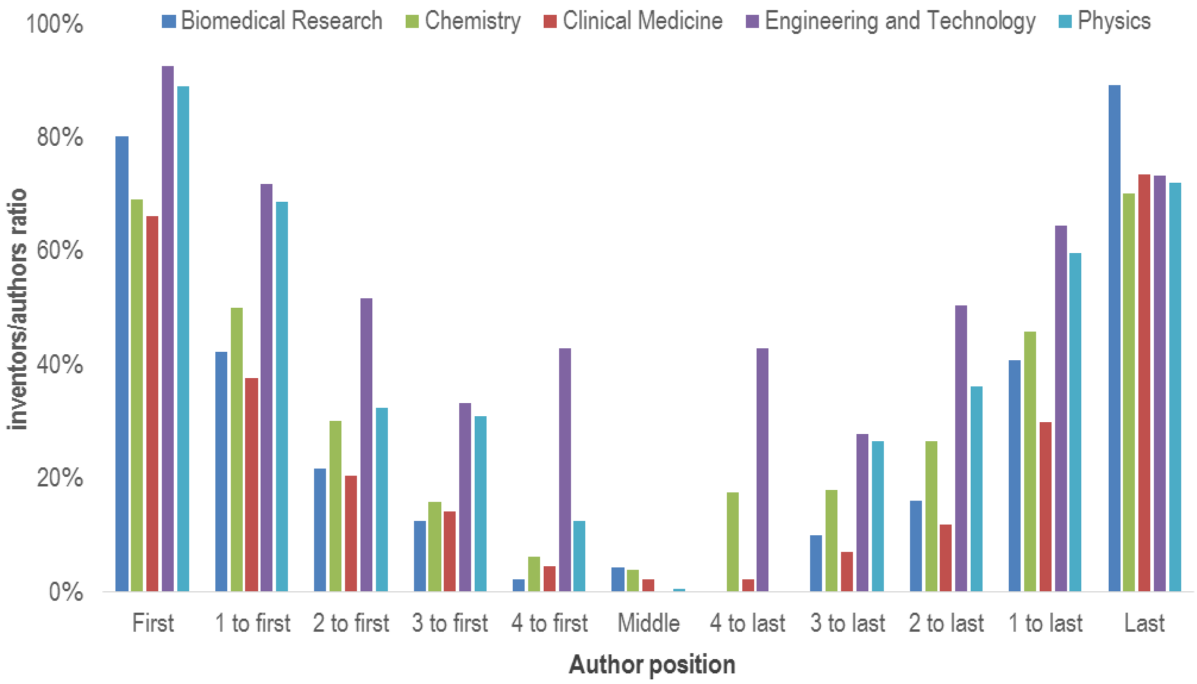


Figure 1. Inventors/authors ratio for first, middle and last authors

As we have previously discussed in previous research (Mongeon, Smith, Joyal, & Larivière, 2016), the first author/middle author(s)/last author division of the byline does not adequately reflect the division of tasks among team members. The authors who are closer to the first and last position have typically contributed more substantially to the work than those listed further in the middle. Thus, Figure 2 displays the inventors/authors ratio as a function of the number of position separating authors from the first or last position. For articles with an odd number of authors, the middle author is duplicated and included in both the left and right side of the middle. For example, the third author of an article with five authors is included in both the "2 to first" and "2 to last" groups. The figure shows that the further an author is from the first and last positions, the least likely he or she is to appear as an inventor on the patent. We also see that in Chemistry, Engineering and Technology, and Physics the inventors/authors ratio tends to remain higher than in Biomedical Research and Clinical Medicine as we move further away from the first and last positions.



Note: the "middle" category includes all authors who are five positions or more (up to 14) away from the first or last position.

Figure 2. Inventors/authors ratio as a function of the number of position separating authors from the first or last position in the byline.

Discussion and conclusion

Our preliminary results display both similarities and differences between disciplines regarding the relationship between authorship and inventorship. Like Ducor (2000), we find that the first and last authors are more likely to appear in the inventors list in all discipline. This is in line with the previous finding that the first and last positions are the most important, and that middle authors have most likely made less substantial or technical contributions (Pontille, 2006).

In Biomedical Research and Clinical Medicine, the last authors are more often inventors than the first authors. Since the last author position is a senior scientist with a supervisory role and the first author is usually a junior scientist leading the experimental work (Pontille 2006), this seems to confirm Haeussler and Sauermann's (2013) finding that hierarchical position predicts inventorship. Interestingly, we find the opposite trend in Engineering and Technology and in Physics. This suggests either that 1) hierarchical position is a weaker predictor of inventorship in these fields, 2) that the author order in these field does not strictly follow the pattern described by Pontille (2006), or 3) that different patterns of division of labor are at play. This last hypothesis is also supported by the fact that middle authors are also more likely to be inventors in Engineering and Technology and Physics than in Biomedical Research and Clinical

Medicine. Chemistry appears to be somewhat in the middle, since the inventors/authors ratio is slightly higher for last authors compared to first authors. This ratio is also higher for middle authors in Chemistry than in Biomedical Research and Clinical Medicine, but lower than in Engineering and Technology and Physics. Further work will be needed to better understand these disciplinary differences.

Further developments of this research include the extension of the analysis to include data over a longer period. This will provide insights on the evolution over time of the authorship and inventorship practices across disciplines. The number of PPPs analyzed might also be further increased by manually identifying valid pairs for which the similarity score is below 0.40. Furthermore, the next step of the study will include contribution data when available, which will allow the comparison of the reported contribution with the information provided by the inventors list. Finally, researchers' characteristics, such as the academic age, prominence (scientific output and impact), and gender will be included in the analysis to measure their potential effect on the patterns observed.

Acknowledgments

The author would like to thank Nees Jan van Eck from the Centre for Science and Technology Studies (CWTS) for the extraction of noun phrases from patents and publications. This research was funded by the Social Sciences and Humanities Research Council of Canada and by the Canada Research Chair on the Transformations of Scholarly Communication.

References

Bennett, D. M. D., & Taylor, D. M. D. (2003). Unethical practices in authorship of scientific papers. *Emergency Medicine*, *15*(3), 263–270.

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, *57*(13), 1758–1770. https://doi.org/10.1002/asi.20380

Cronin, B. (1996). Rates of return to citation. *Journal of Documentation*, *52*(2), 188–197. https://doi.org/10.1108/eb026967

Ducor, P. (2000). Intellectual property: Coauthorship and coinventorship. *Science*, *289*(5481), 873–875. https://doi.org/10.1126/science.289.5481.873

Haeussler, C., & Sauermann, H. (2013). Credit where credit is due? The impact of project contributions and social factors on authorship and inventorship. *Research Policy*, *42*(3), 688–703. https://doi.org/10.1016/j.respol.2012.09.009

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. https://doi.org/10.1177/0306312716650046

Magerman, T., Van Looy, B., & Debackere, K. (2015). Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology. *Research Policy*, *44*(9), 1702–1713. https://doi.org/10.1016/j.respol.2015.06.005

Maienschein, J. (1993). Why collaborate? *Journal of the History of Biology*, *26*(2), 167–183. https://doi.org/10.1007/BF01061964

Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review*, *22*(6), 635. https://doi.org/10.2307/2089193

Meyer, M. (2003). Academic patents as an indicator of useful research? A new approach to measure academic inventiveness. *Research Evaluation*, *12*(1), 17–27. https://doi.org/10.3152/147154403781776735

Mongeon, P., Smith, E., Joyal, B., & Larivière, V. (2016). Partial alphabetical authorship in medical research: an exploratory analysis. In I. Rafols, J. Mollas-Gallart, E. Castro-

Martinez, & R. Wooley (Eds.), *STI2016 - Proceedings of the 21ST international conference on science and technology indicators* (pp. 448–452). Valencia, Spain, Spain.
Packer, K., & Webster, A. (1996). Patenting culture in science: Reinventing the scientific wheel of credibility. *Science, Technology & Human Values*, *21*(4), 427–453. https://doi.org/10.1177/016224399602100403
Pontille, D. (2006). Qu'est-ce qu'un auteur scientifique. *Sciences de La Société*, *67*, 77–93.
Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails: A proposal to make contributors accountable. *JAMA*, *278*(7), 579. https://doi.org/10.1001/jama.1997.03550070071041
Shapiro, D. W., Wenger, N. S., & Shapiro, M. F. (1994). The contributions of authors to multiauthored biomedical research papers. *JAMA*, *271*(6), 438–442. https://doi.org/10.1001/jama.1994.03510300044036
Smith, C. G. (1971). Scientific Performance and the Composition of Research Teams. *Administrative Science Quarterly*, *16*(4), 486. https://doi.org/10.2307/2391768
Wuchty, S., Jones, B. B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099