Type of submission: Presentation

The contribution of middle authors to the production of knowledge in the biomedical field

Philippe Mongeon, Elise Smith, Bruno Joyal, and Vincent Larivière

Introduction

The production of scientific knowledge is increasingly collaborative, a trend is reflected in the increasingly long list of names found in the byline of scholarly articles (Wuchty, Jones, & Uzzi, 2007). Lengthy bylines raise questions as to how the credit and the responsibility should be distributed among the authors (Rennie & Yank, 1998). In the biomedical field, the first and last position are typically given the most importance, the first author often being a junior researcher who lead the research, and the last author being the lab director (Pontille, 2004). In between these two poles are listed an increasingly large number of 'middle authors' who have typically played a more minor role in the research (Larivière et al., 2016). This division of the byline is problematic because it artificially creates a heterogeneous group of middle authors. Indeed, since in large collaborations, there may be different researchers in charge of different parts of the experiment, and more than one lab director, it would seem more adequate to divide the byline three groups containing multiple authors (i.e., first authors, middle authors and last authors). The difficulty is then to decide where to draw the lines delimiting each group. As Zuckerman (Zuckerman, 1968) pointed out, listing only a subset of authors in alphabetical order creates a clear distinction between the authors who are listed alphabetically and the others. This study thus uses partial alphabetical order as a mean to identify middle authors. It aims to provide a better portrait of the division of labor in biomedical research by analysing the prevalence of partial alphabetical order as well as the length of the alphabetical ordered sequences.

We retrieved from the Web of Science all biomedical articles published between 1980 and 2015, and authored by 1 to 100 individuals. Like Waltman (Waltman, 2012), we detected sequences of authors in alphabetical order by giving to each author of a byline an alphabetical rank based on their last name, and then their initials. Then, we corrected for potential errors linked to special character conversion, compound names, name prefixes, indexation errors, and human errors. This was done by concatenating consecutive sequences of three or more alphabetically ordered authors where only one author breaks the alphabetical order, and where the following condition was met

$$(X \leq Y_1 \leq Z)\ or\ (X \leq Y_2 \leq Z)\ or\ (X \leq Y_3 \leq Z)$$

Where *X* is the first letter of the author name before the one causing the break, Y_1 is the first letter of the author name causing the break, Y_2 is the first letter of the author name causing the break after removing potential prefixes, Y_3 is the last initial of the author name causing the break, and *Z* is the first letter of the author name after the one causing the break.

To distinguish incidental and intentional alphabetical order, we defined $b_n(r) = n! - s_n(r)$ as the number of permutations of *n* authors which do *not* contain an alphabetically ordered subsequence of *r* authors. We then, computed the probability of finding an alphabetically ordered subsequence of *r* authors in a list of *n* authors using the following formula. For the present study we set a minimum threshold of r so that the probability of false positive lower than 5%.

$$B_r(X) = \left(\sum_{n\geq 0} \frac{X^{nr}}{(nr)!} - \frac{X^{nr+1}}{(nr+1)!}\right)^{-1}$$

Result

Figure 1 (left panel) shows that the prevalence of alphabetically ordered middle authors correlates with the number of authors on the byline of the articles, while the panel on the right shows that overall, the prevalence of middle authors has been slightly increasing overtime.

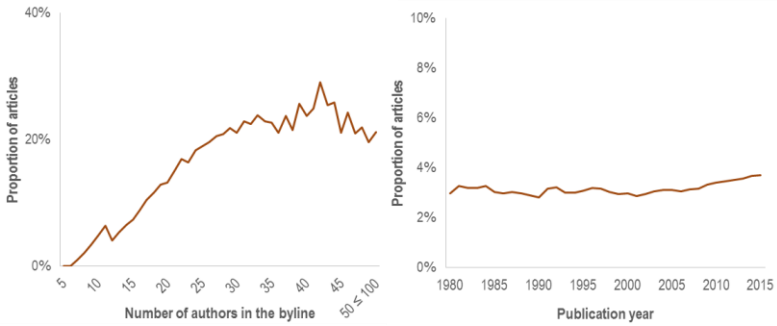

Figure 1. Proportion of article bylines containing different types of alphabetical order as a function of the number of authors (left), and over time (right).

Figure 2 displays the relative share of authorships attributed to first author(s), middle authors, and last authors as a function of the total number of authors (left panel), and as a function of the publication year (right panel). The share of authorships attributed to is generally higher than the share of the first and last authors. The proportion of middle authors slightly decreased with time and as a function of the total number of authors on the byline and, inversely, the share of first and last authors slightly increased.

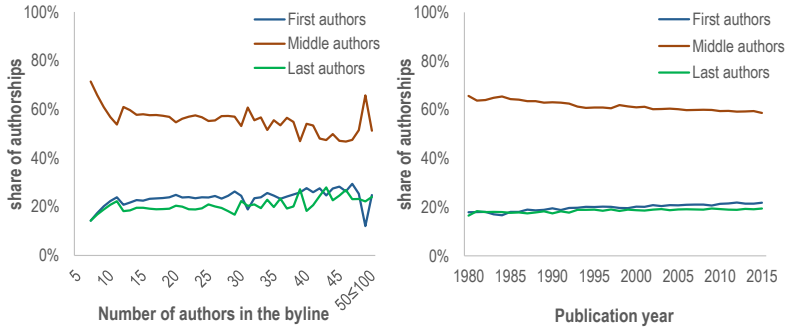

Figure 2. Proportion of authors in the first, middle and last authors groups as a function of the number of authors (left), and over time (right).

Figure 3 shows an almost tenfold increase of the share of total authorships attributed to middle authors over the 1980-2015 period , a perhaps logical consequence of the ever increasing length of the bylines, and the positive correlation between the number of authors and the prevalence of partial alphabetical order. This is clear evidence of the rise of the middle authors within the reward system of science.

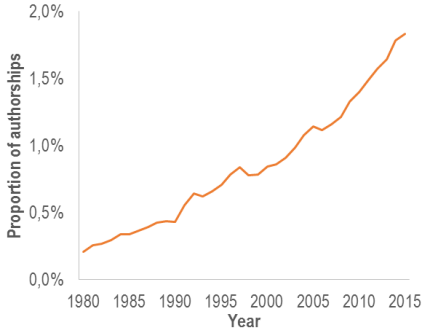

Figure 4. Evolution of the total share of authorships attributed to middle authors on all articles in the biomedical field (1980-2015)

Discussion and conclusion

The results shed new light on the mechanisms that underlie the growing length of the bylines, showing that this trend might be in large part due to an increased division of labor, and consequently to the increased inclusion of individuals with small contributions in the middle of the authors list. Also suggesting a link between the increased attribution of authorship to minor contributors and the longer bylines is that using partial alphabetical order reduces the incentive to keep the author list short. Indeed, when authors are ordered according to their contribution or in full alphabetical order, adding a name on the byline reduces the credit received by each author. Partial alphabetical authorship reduces to a certain extent this "loss of credit" as the different groups of authorship provide a clear delimitation between the main contributors and the peripheral contributors. One could then virtually add to the byline an unlimited number of authors without depriving main contributors of their due credit. It thus appears that not only is name ordering is determined by authorship criteria, but also, inversely, that the name ordering practices partly determines what is (or is not) an author.

Guidelines such as those developed by the International Committee of Medical Journal Editors (ICMJE) clearly define authors as individuals who have "substantially contributed" to the research (International Committee of Medical Journal Editors, 2015). However, when very large numbers of authors are listed on a byline, it seems likely that not all authors meet this criterion. This lack of adherence of the scientific community to the guidelines might indicate that they are somewhat disconnected from the reality of the practices they wish to normalize. Furthermore, widely used indicators like the infamous H-index do not take into account one's position in the byline and then, since minor contributions generally require less work than major contributions, being a middle author may be paradoxically more rewarding than first authorship from a cost-benefit point of view. Inversely, indicators and evaluation processes that put the emphasis on the first position may create disputes and hinder collaboration and the division of labor.

References

International Committee of Medical Journal Editors. (2015). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals.

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. doi:10.1177/0306312716650046

Pontille, D. (2004). *La signature scientifique : une sociologie pragmatique de l'attribution*. Paris: CNRS.

Rennie, D., & Yank, V. (1998). If authors became contributors, everyone would gain, especially the reader. *American Journal of Public Health*, *88*(5), 828–830.

Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, *6*(4), 700–711. doi:10.1016/j.joi.2012.07.008

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, *316*(5827), 1036–1039. doi:10.1126/science.1136099

Zuckerman, H. A. (1968). Patterns of Name Ordering Among Authors of Scientific Papers: A Study of Social Symbolism and Its Ambiguity. *American Journal of Sociology*, *74*(3), 276–291. doi:10.2307/2775535