Alphabetical co-authorship in the social sciences and humanities: evidence from a comprehensive local database[1]

Raf Guns[*]

[*] raf.guns@uantwerpen.be
University of Antwerp, Centre for R&D Monitoring (ECOOM), Middelheimlaan 1, 2020 Antwerp, Belgium

Introduction

A recurring debate related to research evaluation focuses on the question how multi-authored publications should be counted. For the sake of simplicity, we assume that one (single-authored) publication counts for one credit, but the argument is the same if the credit for a publication can vary based on other factors, such as citation count.

Two classic strategies for credit assignment are whole counting and fractionalized counting. Whole counting implies that each author receives one full credit for the publication, regardless of the number of authors. Fractionalized counting occurs when the credit is divided among the authors. In the simplest case, each of n authors receives $1/n$ credit.

More involved fractionalization strategies try to account for the relative contribution of each author, such that the author who has contributed the most to the paper receives the most credit, and so on. Most prior research starts from the assumption that the first author is the most important one, followed by the second one, and so on. Possible approaches include: only the first author receives credit (Cole & Cole, 1973; Persson, 2001), geometric counting (Egghe, Rousseau, & Van Hooydonk, 2000), and harmonic credit allocation (Hagen, 2008, 2013). Sometimes, other ways of determining author importance are used, such as recognition of the corresponding author or author contribution statements. Some research has pointed to the importance of the last author (Zuckerman, 1968) or the rise of 'equal first authors' (Hu, 2009).

Use of the byline order implies that authors deliberately order the byline according to relative contribution. A different criterion according to which authors can be ordered in the byline is simply by alphabetical order of their names. In that case, order of authors in the byline should not be used for credit allocation (Van Praag & Van Praag, 2008). Previous research (Frandsen & Nicolaisen, 2010; Levitt & Thelwall, 2013; Waltman, 2012) has already looked into the question to what extent alphabetical co-authorship occurs through time and in different disciplines. Major findings of these studies include:
- The number of authors per publication has increased over the last few decades;
- Some disciplines, like economics and mathematics, have a strong culture of alphabetical co-authorship whereas others do not;
- Overall, alphabetical authorship is declining.

These studies typically include (parts of) the social sciences and humanities or SSH in short. However, there are at least two reasons that suggest that, at this moment, no clear picture is available for the whole of SSH. First, most studies are concerned only with (parts of) the social sciences and much less with the humanities. Second, these studies are typically based on large international databases such as Web of Science (WoS). Since these databases only cover a limited subset of publications from SSH (Archambault et al., 2013), it is as of yet unknown to what extent alphabetical co-authorship is used in in SSH.

In this paper, we try to partially alleviate these limitations by using a comprehensive local database as data source. Specifically, we want to address the following questions:
1. To what extent do co-authored publications in SSH use alphabetical co-authorship?
2. How do different SSH disciplines differ in their use of alphabetical co-authorship?
3. How does alphabetical co-authorship vary with publication type?
4. How has alphabetical co-authorship evolved over the past decade?

Data and methods

Data

This paper uses data from the Flemish Academic Bibliographic Database for the Social Sciences and Humanities or VABB-SHW (www.ecoom.be/en/vabb). This is a comprehensive database of all peer reviewed publications by SSH researchers affiliated to a Flemish university from the year 2000 onwards (Engels, Ossenblok, & Spruyt, 2012; Ossenblok, Verleysen, & Engels, 2014).

The full data set covers the time period 2000–2013 and consists of bibliographic information for 59,560 peer reviewed publications with at least one author/editor affiliated to an SSH department at a Flemish university. There are 34,683 publications (journal articles, monographs, book chapters, or proceedings papers) with more than one author and 1,210 publications (edited books) with more than one editor. This subset of 35,893 publications (60.3% of all publications) is the basis of the present paper.
Table 1 provides additional details on the publication types and their counts in the data. Each publication is also assigned to one or more disciplines, according to the departmental affiliation of its authors; there are 9 disciplines in the humanities and 7 disciplines in the social sciences. In addition to those, the VABB-SHW has the general disciplines 'Humanities general' and 'Social sciences general'.

	1 author/editor	>1 author/editor	Total
Journal articles	16,801	29,208	46,009
Monographs	621	287	908
Edited books	407	1,210	1,617
Book chapters	5,326	3,624	8950
Proceedings papers	512	1,564	2,076
Total	**23,667**	**35,893**	**59,560**

Table 1. Overview of publication types in data set

Publications in the VABB-SHW belong to two separate subcategories: VABB-WoS publications are publications that are also indexed in one of the WoS databases, whereas VABB-GP publications are not in WoS. The latter have been selected by the Authoritative Panel (Gezaghebbend Panel or GP), an independent body of academics charged by the Flemish Government with upholding the criteria for inclusion in the VABB-SHW, such as peer review. The majority of publications with more than one author/editor are VABB-WoS publications ($n = 20,298$ or 56.6%).

Methods

We can distinguish between intentional and accidental alphabetical co-authorship. Intentional alphabetical co-authorship (IAC) occurs when the authors have made a deliberate choice to put their names in alphabetical order on the publication, whereas accidental alphabetical co-authorship (AAC) occurs when the names are in alphabetical order but the byline order has been established using some other criterion (e.g., respective contribution to the end result).

First, we normalize names to account for slight variations in the way names are written. Spaces and other non-letter characters in author names are omitted, as well as accents. Full first names are reduced to the first letter. A name like 'De Pré, Johan', for instance, is normalized to 'DEPRE, J'.

Next, we determine the number and proportion of publications (for the whole as well as per subgroup, e.g. per discipline or publication type) that have two or more authors and that are alphabetically authored. The resulting number and proportion count both IAC and AAC, and hence constitute an overestimation of the proportion of IAC. Other things being equal, for a publication with two authors that uses another criterion for the order in the byline, there is a 50% chance that the end result will be alphabetical and we end up with AAC. The probability of AAC rapidly decreases as the number of authors increases. In general, for a publication with n authors, the probability of AAC equals $1/n!$. Since we cannot directly measure IAC, we need to estimate it, accounting for the fact that, e.g., a paper with five authors in alphabetical order is much more likely due to IAC (99%) than a paper with only two (50%).

Waltman (2012) provides an unbiased estimator p for the proportion of IAC in a body of N publications, which is in turn based on a model by Van Praag & Van Praag (2008):

$$p = \frac{\sum_{i=1}^{N}\left(\frac{a_i - 1/n_i!}{1 - 1/n_i!}\right)}{N} \qquad (1)$$

where n_i denotes the number of authors of publication i, and a_i is 1 if publication i is alphabetically authored and 0 otherwise. Note that p can only be applied to those publications that have more than one author. The estimator p typically ranges between 0 (no IAC) and 1 (complete IAC), although theoretically it can also yield negative values. A negative value

would imply that authors deliberately seek out non-alphabetical orderings. Using formula (1), we can estimate the percentage of IAC in a body of co-authored publications, controlling for the number of authors on each individual publication.

Results

Table 2 summarizes the results per discipline. All numbers reported here, as well as in the following tables, are based on the subset of publications with two or more authors or editors. Hence, the mean and median number of authors reported are higher than what has been reported in the literature, where single-authored publications are also counted (Ossenblok & Engels, 2015; Ossenblok et al., 2014).

We find large disciplinary differences in the proportion of alphabetical co-authorship. The proportion of alphabetical co-authorship ranges from 6.9% for Social health sciences to 48.1% for Literature. Although this is partially related to the number of co-authors per paper – the median for Social health sciences is 5, whereas for Literature this is only 2 –, the proportion of IAC for these two disciplines is still respectively lowest and highest. Remarkably, IAC is virtually non-existent in Social health sciences. The second and third places for highest proportion of IAC go to, respectively, Economics & business and History. Overall, we find that alphabetical co-authorship and IAC occur in almost every discipline, but in none of the disciplines are they the default or most used option. As can be seen from Table 2, alphabetical co-authorship and, especially, IAC occur more in the humanities than in the social sciences. While the humanities as a whole exhibit an IAC proportion of 12.0%, this is only 7.3% for the social sciences. Both percentages are, however, still significantly higher than the overall 3.7% reported by Waltman (2012), indicating that alphabetical co-authorship is more engrained in SSH than in science as a whole.

Table 2. Overview of results per discipline

	N	Mean authors	Median authors	Alphabetical order	IAC
Literature	563	2.89	2	48.1%	23.3%
Economics & business	7485	3.54	3	36.0%	18.7%
History	764	3.23	2	41.4%	16.3%
Criminology	1110	3.31	2	39.0%	14.9%
Philosophy	1482	3.89	3	34.7%	13.7%
Linguistics	2341	3.41	3	35.5%	12.5%
Theology	401	3.72	3	33.2%	11.8%
Humanities general	1434	3.43	3	33.6%	11.2%
Law	2508	2.82	2	42.2%	10.9%
Political sciences	1463	2.89	2	38.8%	10.4%
Communication studies	1168	3.71	3	32.1%	9.9%
Sociology	2228	4.32	3	24.5%	6.8%
History of arts	871	4.01	4	24.2%	6.5%
Archaeology	599	5.03	4	17.4%	6.2%
Social sciences general	5794	4.89	4	16.7%	3.8%
Educational sciences	2327	3.84	3	20.6%	3.2%
Psychology	5638	4.71	4	15.7%	1.9%
Social health sciences	8023	6.65	5	6.9%	0.4%

Next, we turn to the question to what extent alphabetical co-authorship varies with publication type. The results are summarized in Table 3. Monographs and edited books are quite different from the other three publication types, in that they display a much larger share of IAC. The high percentage of IAC for edited books confirms previous research (Ossenblok & Engels, 2015; Ossenblok, Guns, & Thelwall, 2015) that edited books and editors are quite different from other publication types and their authors. The low proportion of IAC for proceedings papers is somewhat unexpected and deserves further investigation. Journal articles and book chapters exhibit a comparable proportion of alphabetical co-authorship and IAC.

Table 3. Overview of results per publication type

	N	Mean authors	Median authors	Alphabetical order	IAC
Journal articles	29,195	4.53	3	24.2%	7.6%
Monographs	284	3.25	3	43.3%	21.4%
Edited books	1,209	3.08	3	49.0%	30.0%
Book chapters	3,623	3.59	3	29.5%	9.2%
Proceedings papers	1,564	4.66	4	15.0%	2.6%

Given the large amount of journal articles, the results in Table 3 cannot show all variation that exists within this publication type. We therefore look into the question to what extent alphabetical co-authorship varies with whether or not a journal is indexed in WoS. As was mentioned earlier, the majority of co-authored publications are VABB-WoS publications. Indeed, we know from prior research that more internationally oriented literature – often indexed in databases like WoS or Scopus, authored in English, etc. – tends to have more authors (Ossenblok et al., 2014). This can also be seen from the mean and median number of

authors for VABB-WoS compared to VABB-GP journal articles (Table 4). The difference between WoS-indexed and non-WoS-indexed publications is even more striking when we consider byline order. Both alphabetical co-authorship and IAC occur about twice as frequently in publications that are not indexed in WoS. This discrepancy can be seen in both the humanities and social sciences, although the difference is most pronounced for the social sciences. This strongly suggests that alphabetical co-authorship within SSH is most prevalent in locally or regionally oriented literature, even if we account for the fact that the number of co-authors is typically lower there.

Table 4. Comparison of VABB-WoS and VABB-GP journal articles

		N	Mean authors	Median authors	Alphabetical order	IAC
VABB-GP	Humanities	3,996	2.83	2	42.0%	11.8%
	Social sciences	6,732	3.47	3	32.4%	10.8%
	All	9,834	3.22	3	35.8%	10.8%
VABB-WoS	Humanities	2,759	4.31	3	26.2%	7.2%
	Social sciences	17,423	5.35	4	17.0%	5.6%
	All	19,361	5.19	4	18.4%	5.9%

Figure 1 displays the evolution of alphabetical co-authorship and IAC between 2000 and 2013. After a relatively strong decline over the first five years, the proportions remain fairly stable around 7 to 8%. This is broadly in line with Waltman's (2012) finding that alphabetical co-authorship is declining.

Figure 1. Evolution of proportions of alphabetical order and IAC (2000–2013)
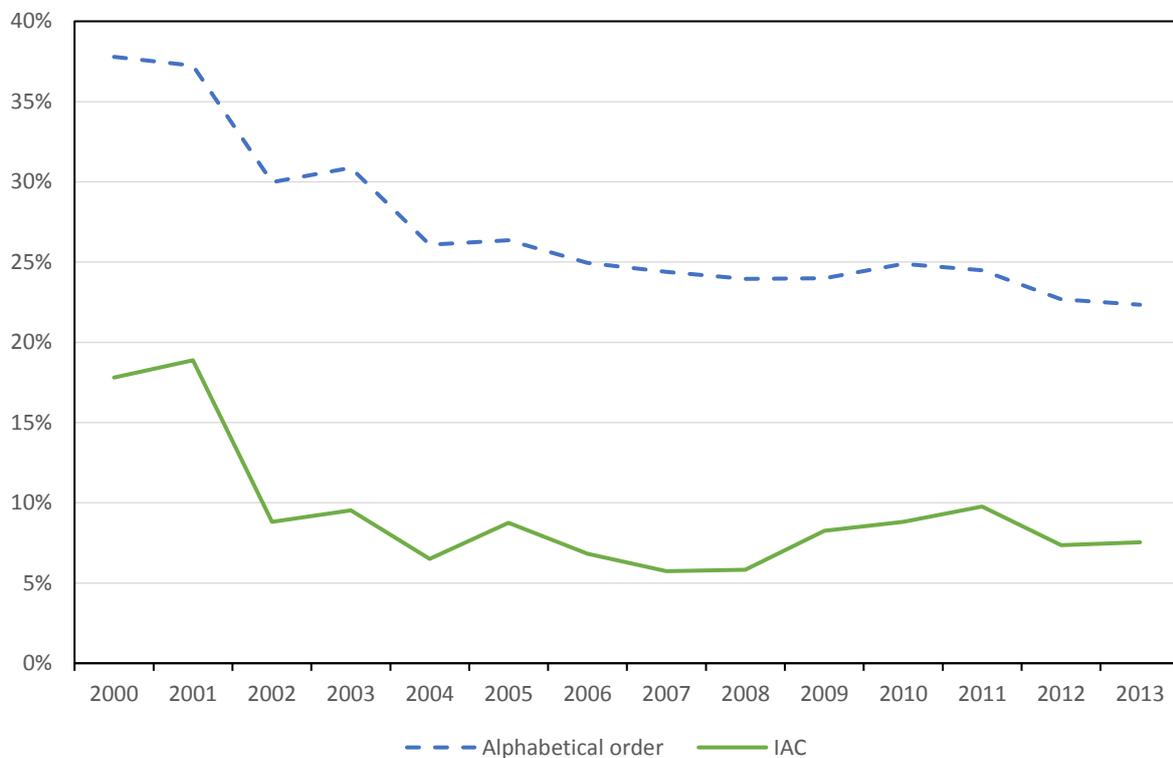
Discussion and conclusion

We have shown that alphabetical co-authorship occurs more in SSH than in other fields of science and that authors in the humanities choose an alphabetical ordering of their names in the byline more frequently than in the social sciences. At the same time, these overall findings should not obscure the fact that there exists quite a bit of variety within the social sciences and humanities. The highest proportions of IAC are found for Literature, Economics & business, and History. The use of alphabetical co-authorship is declining over time.

The proportions of alphabetical co-authorship and IAC in Economics & business we find are clearly lower than those reported in previous research (Frandsen & Nicolaisen, 2010; Levitt & Thelwall, 2013; Waltman, 2012), where, for instance, the proportion of alphabetical co-authorship was roughly around 75%. The main explanation lies in the way publications are assigned to disciplines in the VABB-SHW. A publication is assigned to a discipline if at least one of its authors belongs to an administrative unit that is assigned to this discipline. In the case of Economics & business this has led to a situation where many publications are only tangentially related to the core fields of Economics & business. This was verified by looking at the subset of WoS publications from this discipline. If we restrict the WoS publications from Economics & business to those with WoS SC Economics (20.3%), we find an IAC proportion of 48.4%. However, this is still lower than the 58% reported by Waltman (2012).

As for publication types, alphabetical co-authorship occurs most often in books, is less common in articles in journals or in books, and is rare in proceedings papers. Articles in WoS journals exhibit less alphabetical co-authorship than those in GP-selected journals. The differences between publication types cannot be explained solely through disciplinary preferences for certain publication types (Nederhof, 2006), given the high percentage of IAC for edited books and monographs.

Ideally, the order of authors in the byline should not be used for credit assignment in those disciplines where the proportion of IAC exceeds a certain threshold. Even if the proportion of IAC stays well below the threshold, this does not necessarily imply that the byline order accurately reflects each author's contribution.

References

Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2013). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, *68*(3), 329–342.

Cole, J. R., & Cole, S. (1973). *Social stratification in science*. Chicago: U.Chicago Press.

Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the American Society for Information Science*, *51*(2), 145–157.

Engels, T. C. E., Ossenblok, T. L. B., & Spruyt, E. H. J. (2012). Changing publication patterns in the Social Sciences and Humanities, 2000–2009. *Scientometrics*, *93*(2), 373–390.

Frandsen, T. F., & Nicolaisen, J. (2010). What is in a name? Credit assignment practices in different disciplines. *Journal of Informetrics*, *4*(4), 608–617.

Hagen, N. T. (2008). Harmonic allocation of authorship credit: source-level correction of bibliometric bias assures accurate publication and citation analysis. *PloS One*, *3*(12), e4021.

Hagen, N. T. (2013). Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy. *Journal of Informetrics*, *7*(4), 784–791.

Hu, X. (2009). Loads of special authorship functions: Linear growth in the percentage of "equal first authors" and corresponding authors. *Journal of the American Society for Information Science and Technology*, *60*(11), 2378–2381.

Levitt, J. M., & Thelwall, M. (2013). Alphabetization and the skewing of first authorship towards last names early in the alphabet. *Journal of Informetrics*, *7*(3), 575–582.

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A Review. *Scientometrics*, *66*(1), 81–100.

Ossenblok, T. L. B., & Engels, T. C. E. (2015). Edited books in the Social Sciences and Humanities: Characteristics and collaboration analysis. *Scientometrics*, *104*(1), 219–237.

Ossenblok, T. L. B., Guns, R., & Thelwall, M. (2015). Book editors in the social sciences and humanities: an analysis of publication and collaboration patterns of established researchers in Flanders. *Learned Publishing*, *28*(4), 261–273.

Ossenblok, T. L. B., Verleysen, F. T., & Engels, T. C. E. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities (2000-2010). *Journal of the Association for Information Science and Technology*, *65*(5), 882–897.

Persson, O. (2001). All author citations versus first author citations. *Scientometrics*, *50*(2), 339–344.

Van Praag, C. M., & Van Praag, B. M. S. (2008). The Benefits of Being Economics Professor A (rather than Z). *Economica*, *75*(300), 782–796.

Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, *6*(4), 700–711.

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.