

Is Paper Uncitedness a Function of the Alphabet?

Clément Arsenault¹ and Vincent Larivière²

¹ *clement.arsenault@umontreal.ca*

École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal,
Montréal (Qc) (Canada)

² *vincent.lariviere@umontreal.ca*

École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal,
Montréal (Qc) (Canada), and

Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la
Technologie (CIRST), Université du Québec à Montréal, Montréal (Qc) (Canada).

Introduction

Citation counts are well-established measures of researchers' scientific impact. One would assume that external factors, such as someone's name, over which an individual has little control over, does not influence such indicators. Yet, reference lists and—to a lesser extent—search results from online databases, are often presented in alphabetical order sorted by first author surname. A large number of scientific journals use parenthetical referencing styles (a.k.a. Harvard referencing style) in which partial parenthetical citations (such as author+date or author+title) are embedded in the text, accompanied by an alphabetized list of complete citations at the end. These lists may be consulted to locate a specific item (known-item search) but are also used in a scanning mode, usually from top (A) to bottom (Z), to identify papers that would potentially provide answers to a question or reinforce an argument.

In marketing and advertising research it is well recognized that product positioning influences choice and selection and that usually “first is best”, i.e., that items presented first usually have a better chance of being selected (Carney & Banaji, 2012). Such a phenomenon has also been observed by Haque and Ginsparg (2009, p. 2215) who measured a significant correlation between article position in the arXiv repository and citation impact, due the “visibility” effect that “can drive early readership, with consequent early citation potentially initiating a feedback loop to more readership and citation.” Order of presentation (or scanning order) is also central to Cooper's utility theory (1971) since items consulted earlier will find a better chance of being useful to a searcher.

Taking these elements into account, authors with a surname whose initial letter arrives early in the alphabet get more visibility, a situation that is further compounded by the fact that in multi-authored papers, authorship order is sometimes determined by alphabetical rank. This practice is even fairly common in some fields such as economics and finance, mathematics, high-energy physics, market-

ing, political science, international relations and law (Frandsen & Nicolaisen, 2010, p. 615; Levitt & Thelwall, 2012, p. 725; Waltman, 2012, p. 701). In the field of economics where authorship order is almost always determined alphabetically, research has shown that economists with early surnames (i.e., with initial letters that occur early in the alphabet) publish more articles (van Praag & van Praag, 2008), are more likely to get employment at high standard research departments (Efthyvoulou, 2008) and receive more tenure at top economic departments (Einav & Yariv, 2006), since “the order of authorship, rather than contributorship, is commonly used to assess the prestige that an author incurs from a published research study” (Chambers, Boath, & Chambers, 2001, p. 1461).

Literature Review

Citation likelihood based on author's surname position in the alphabet has also been the subject of some recent studies. McCarl (1993) found that authors receive approximately 0.5% less first author citations per letter the latter their names are in the alphabet. Laband and Tollison (2006) showed that “alphabetized co-authored papers with two authors are more highly cited than non-alphabetized co-authored papers” in both economics and agricultural economics. In a large-scale study Huang (2015, p. 780) revealed that “papers with first authors whose surname initials appear earlier in the alphabet get more citations [and that this effect] is significantly stronger in those fields with longer reference lists.”

This later observation reinforces the idea that the browsing effect is to the advantage of papers listed towards the top of alphabetized reference lists since readers are more likely to run out of patience before they get to the end of the list. To corroborate these findings, our study will look at the reverse effect, namely the greater invisibility of papers appearing at the end of reference lists by measuring the uncitedness rates of papers as correlated to the first author's position in the alphabet.

Data and Methodology

The data set used in this study was obtained from the Web of Science databases and consists of all the scientific papers published between the years 2000 and 2013, totalling 15,056,841 source items. Papers are assigned to one of the fourteen disciplines of the National Science Foundation (NSF) classification. Field-normalized citations rates for each paper were calculated, and grouped by the first letter of the surname of the first author, which means that each paper was counted only once in the dataset.

Results and Discussion

Preliminary analysis reveals that, in most of the fourteen NSF disciplines, uncitedness rates tend to increase with the progression of the first author's last name in the alphabet indicating that papers with a first author whose last name starts with a letter that occurs later in the alphabet might be less visible. Correlation coefficients are the strongest in the disciplines of Mathematics and Physics (figure 1) indicating that the practice in these disciplines to list co-authors on the basis of author's position in the alphabet seems to exacerbate this problem.

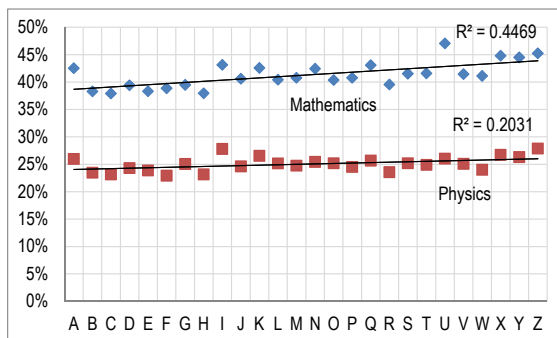


Figure 1. Uncitedness rates of Mathematics and Physics papers by initial letter of first author's surname.

Further analysis at the level of specialty of the NSF classification will validate whether such effects are observable in other fields (such as Economics & Finance) where the tradition of listing co-authors alphabetically is highly prevalent, as well as the potential effect of researchers from specific countries whose surnames are more likely to start with a letter that appear towards the end of the alphabet. On the whole, these results show that papers whose first author bears a surname that is at the end of the alphabet are at a disadvantage in terms of citation rates, a finding that is likely a consequence of the current structure of reference lists and of search results from online databases.

In a more detailed analysis, confounding factors such as the higher prevalence of names beginning with some letters and the concentration of names from certain regions will be considered.

References

- Carney, D. R., & Banaji, M. R. (2012). First is best. *PLoS ONE*, 7(6), e35088. doi:10.1371/journal.pone.0035088
- Chambers, R., Boath, E., & Chambers, S. (2001). The A to Z of authorship: Analysis of influence of initial letter of surname on order of authorship. *BMJ*, 323(22–29 Dec.), 1460–1461. doi:10.1136/bmj.323.7327.1460
- Cooper, W. S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval*, 7(1), 19–37. doi:10.1016/0020-0271(71)90024-6
- Efthymiou, G. (2008). Alphabet economics: The link between names and reputation. *The Journal of Socio-Economics*, 37(3), 1266–1285. doi:10.1016/j.socec.2007.12.005
- Einav, L., & Yariv, L. (2006). What's in a surname?: The effects of surname initials on academic success. *Journal of Economic Perspectives*, 20(1), 175–188. doi:10.1257/089533006776526085
- Frandsen, T. F. & Nicolaisen, J. (2010). What is in a name?: Credit assignment practices in different disciplines. *Journal of Informetrics*, 4(4), 608–617. doi:10.1016/j.joi.2010.06.010
- Haque, A., & Ginsparg, P. (2009). Positional effects on citation and readership in arXiv. *Journal of the American Society for Information Science and Technology*, 60(11), 2203–2218. doi:10.1002/asi.21166
- Hart, R. L. (2000). Co-authorship in the academic library literature: A survey of attitudes and behaviors. *Journal of Academic Librarianship*, 26(5), 339–345. doi:10.1016/S0099-1333(00)00140-3
- Huang, W. (2015). Do ABCs get more citations than XYZs? *Economic Inquiry*, 53(1), 773–789. doi:10.1111/ecin.12125
- Levitt, J. M. & Thelwall, M. (2012). Alphabetical co-authorship in the Social Sciences. *Proceedings of the 17th International Conference on Science and Technology Indicators Conference* (Montreal, Canada). http://2012.sticonference.org/Proceedings/vol2/Levitt_Alphabetical_523.pdf
- McCarl, B. A. (1993). Citations and individuals: First authorship across the alphabet. *Review of Agricultural Economics*, 15(2), 307–312. doi:10.2307/1349450
- 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. doi:10.1111/j.1468-0335.2007.00653.x
- 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