

Alphabetization and the Skewing of First Authorship Towards Last Names Early in the Alphabet¹

Jonathan M. Levitt^{a*} and Mike Thelwall^b

a Department of Information Science, Loughborough University, Loughborough, LE11 3TU, UK and Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK. b Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK. *Corresponding author

The practice of listing co-author surnames in alphabetical order, irrespective of their contribution, can make it difficult to effectively allocate research credit to authors. This article compares the percentages of articles with co-authors in alphabetical order (alphabetization) for two-author, three-author and four-author articles in eighteen social sciences in 1995 and 2010 to assess how widespread this practice is. There is some degree of alphabetization in all disciplines except one but the level varies substantially between disciplines. This level is increasing slightly over time, on average, but it has increased substantially in a few disciplines and decreased in others, showing that the practice of alphabetization is not fading away. A high correlation between alphabetical order and the proportion of first authors near the beginning of the alphabet confirms that high percentages of alphabetical order could affect the appropriate allocation of research credit. Similar patterns were found for science and the humanities. Finally, since some degree of alphabetization is almost universal in social science disciplines, this practice may be affecting careers throughout the social sciences and hence seems indefensible.

1. Introduction

An investigation by Einav and Yariv (2006) found that (a) in 88% of the multi-authored articles published in five prominent economics journals during 1980-2002 the authors were listed alphabetically and (b) tenured faculty at the top 35 economics departments had, in general, surnames higher in the alphabet (e.g., A, B) than did untenured faculty. Those findings suggest that, for disciplines with a high percentage of authors' surnames in alphabetical order, authors with surnames higher in the alphabet might be more likely to be promoted in academia. This is worrying evidence of bias in a single social science discipline. An investigation into the careers of 1500 chemists at British universities has also found a relationship between career advancement and the position of the author's surname in the alphabet (Rudd, 1977), confirming that economics is not an isolated case.

One reason for a possible link between position in alphabet and academic success is that, for disciplines with an abnormally high percentage of author's surnames in alphabetical order (called 'alphabetized' articles), authors with surnames higher in the alphabet may be more likely to be first authors than authors with surnames lower in the alphabet. Hence, disciplines with particularly high percentages of alphabetized articles pose two critical problems for those seeking to allocate credit for research: (a) how to decide whether the author order of alphabetized articles reflects author contributions, and (b) how to avoid favouring authors whose surnames are higher in the alphabet, by giving them the major credit for alphabetized articles in which they did not contribute the most.

Although there is a small amount of research into author alphabetization, no study is comprehensive for an area of science and covers multiple years. The current article fills this gap with an analysis of eighteen social science disciplines. Moreover, no research has

¹ Levitt, J., & Thelwall, M. (2013). Alphabetization and the skewing of first authorship towards last names early in the alphabet. *Journal of Informetrics*, 7(3), 575–582.

investigated the impact of an author's position in the alphabet on their likely authorship positions within multi-authored articles. The current article investigates, for the same social science disciplines, the relationship between the percentage of alphabetized articles and any skewing of first authorship towards researchers with surnames higher in the alphabet.

2. Related literature and research questions

The question of the position of an author's surname in the alphabet was considered in the 1970s (Rudd, 1977) and the question of authorship order was considered in the 1980s (Good, 1989). In the 1990s it was suggested that it was important that the order of authors reflected their contribution to an article (Savitz, 1999). Nevertheless, although it is often perceived that the first named author has made the largest contribution to an article (van Praag & van Praag, 2008) the assignment of credit varies from discipline to discipline (Engers, Gans, Grant & King, 1999) and even within disciplines (Laband, 2002; Joseph, Laband & Patil, 2005). This impacts on some citation counting methods used in research evaluation. In the harmonic counting system, authors higher in the author list are automatically favoured because author credit is inversely proportional to the position of the author in the article (Hodge & Greenberg, 1981; Hagen, 2010). In contrast, the Pareto weighting counting system allocates credit to authors purely on the basis of their citation record (Tol, 2011). Perhaps even more significantly, authorship may sometimes be granted for purely social or political reasons (Davenport & Cronin, 2001), potentially distorting the credit given to the individual contributors. In consequence, some journals now require explicit statements about how much each author has contributed to a published article.

There have been some studies on how researchers decide on authorship order. A review of publication practice claimed that for researchers in many sciences the amount of work was the preferred method for determining authorship order and neither prestige nor position were considered (Marusic, Bosnjak & Jeroncic, 2011) but this may vary amongst disciplines. Nevertheless, an investigation of articles by at least three co-authors in four medical journals found that, in general, the participation by the first author was the most, the last author second largest and the second author third largest (Baerlocher, Newton, Gautam, Tomlinson & Detsky, 2007). This demonstrates that credit allocation is not necessarily straightforward when alphabetization is not used. Perhaps related to this finding, an investigation of co-authorship in Biology and Biomedicine, Materials Science and Natural Resources research in Spain found that younger researchers and those in the lower professional ranks were more likely to be first authors, whereas older or more highly-ranked researchers were more likely to be the last listed author (Costas & Bordons, 2011).

Some research has focused on the extent to which authorship occurs in alphabetical order. High-energy physics generally lists authors in alphabetical order (Birnholtz, 2007) and, as discussed below, alphabetization is present to some extent in many other fields. An investigation of economics journals found that about 70% of the articles published in the top thirty-six economics journals were alphabetized whereas only about 60% of the articles in other economics journals were alphabetized (Hilmer & Hilmer, 2005). Possibly related to this, an investigation of economics and agricultural economics found that in general two-author alphabetized articles were more highly cited than non-alphabetized two-author articles (Laband & Tollison, 2006); however the study also found that alphabetized articles by more than two authors were not more highly cited than non-alphabetized articles by the same number of authors. This suggests that two-author articles may have a different dynamic from that of articles by more than two authors. In a more wide ranging study, Waltman (2012) identified, for articles published in 2011, the 25 Web of Science categories with the highest level of alphabetization, including 1: Mathematics, 2: Business, Finance, 3: Economics, 4:

Philosophy, 5: Social Sciences, Mathematical Methods, 6: Physics, Particles & Fields, 7: Mathematics, Applied, 8: Political Science, 9: International Relations and 10: History.

The level of alphabetization has changed over time. Frandsen and Nicolaisen (2010) found a slight increase in the percentage of alphabetized articles in economics during the period 1988 to 2007, but during that same period a sharp decrease in the percentage in information science. Waltman (2012) investigated the variation over time in the percentage of articles intentionally put in alphabetical order in four categories (Business, Finance; Economics; Mathematics; Physics, Particles & Fields). For each category there was only a limited change between 1995 and 2010. The current study seeks to extend these findings by asking the following question.

Question 1 (Variation in alphabetization over time): *For social science disciplines to what extent does the percentage of alphabetized articles vary from year to year?*

As a secondary issue, this article also identifies alphabetisation on a more detailed level than before. Waltman's 2012 study provides a combined statistic for alphabetization based on a mathematical model, but does not provide separate statistics for publications with different numbers of co-authors; this current study is more fine-grained in that it obtains separate data for different levels of authorship.

Finally, one investigation has studied specific letters of the alphabet. Based on 550 articles and editorials published in the British Medical Journal in 2000 and 2001, Chambers, Boath and Chambers (2001) found that first authors were more frequent than second or third authors for nine of the 13 letters in the first half of the alphabet (A, E, F, G, H, I, J, L and M), but only for two letters in the second half of the alphabet (P and Y). The findings were based on a small sample; only seven author surnames began with a Y. Nevertheless, the study indicates a skewing of first authorship towards authors with surnames higher in the alphabet for one journal. It seems likely that in disciplines with a high percentage of alphabetized articles the percentage of first authorship is generally skewed towards the beginning of the alphabet. In order to test this hypothesis, this study asks the following question.

Question 2 (First author skewing): *What is the relationship between a discipline having a high percentage of alphabetized articles and the skewing of the surnames of first author towards the beginning of the alphabet?*

This is relevant because first author skewing can occur without full alphabetization, for example if the authors are listed in alphabetical order except for a senior researcher being placed at the end of the list. Skewing is an important issue when evaluating researchers; unless skewing is taken into account, researchers with names higher in the alphabet are particularly likely to be given credit for first-authorships which do not reflect their contribution. Moreover, the link between alphabetization and first authorship is not straightforward due to the potential for partial alphabetization.

3. Data and Methods

Data was extracted from the Social Sciences Citation Index (SSCI) on the subject area Psychology and the 17 SSCI categories (apart from Clinical Neurology, Neurosciences, Psychiatry and categories containing the word 'Psychology') with at least 1,500 SSCI articles published in 2003. Clinical Neurology and Neurosciences were excluded as they were regarded as more sciences than social sciences and the year 2003 was chosen as it is roughly midway between 1995 and 2010 (the start and end years of the study). The minimum requirement of 1500 articles was designed to ensure that enough data was available, on each discipline analysed, to give clear results because the individual categories analysed below within smaller fields might be too small to give useful results. The merging of the different psychology classes was necessary because they represented similar subject areas. The

average number of articles per discipline ranged from 3,940 in 1995 to 8,103 in 2010. Only Web of Science documents of type ‘article’ were included.

Variation in alphabetization over time (Question 1) was investigated by calculating, for each discipline, the percentage of alphabetized articles among the articles by two, three and four authors. The percentages of alphabetized articles were compared with the expected percentages if the order of author surnames were to be random (called here ‘expected random value’, or ERV). The ERV is 50% for two-author articles, 16.7% for three-author articles and 4.2% for four-author articles. These three percentages are independent of the distribution of author names in the alphabet.

First author skewing (Question 2) was investigated by examining the relationship between level of alphabetization and the normalised distribution of the first letter of the surnames of the first authors. This distribution was normalised by dividing the frequency of first authorship by the frequency of all authorship and for this reason the findings on first authorship are not dependant on the distribution of all authorship (e.g., whether there are concentrations of author names beginning with particular letters of the alphabet). In order to identify some disciplines in science and the humanities with high levels of skewing, this study also investigates first author skewing for the disciplines identified by Waltman (2012) as having particularly high levels of alphabetization.

On average across the eighteen SSCI disciplines investigated to address both research questions: (a) for 1998, 24% of articles were by two authors, 14% by three authors and 7% by four authors, (b) in 2004, 25% of articles were by two authors, 16% by three authors and 9% by four authors, and (c) for 2010, 25% of articles were by two authors, 18% by three authors and 11% by four authors. Table 1 indicates that the percentage of multi-author articles ranged from 16.3% for International Relations to 74.3% for Business (with an average across disciplines of 50.4%); the percentage of four-author articles ranged from 0.6% for International Relations to 12.4% for Public Health (with an average across disciplines of 5.5%). With the exception of International Relations four-author articles, the number of articles exceeds twenty in each discipline and level of authorship in the table. In this table and all other tables in this study, the mean is calculated by weighting each discipline equally.

Table 1: Percentages of articles at different levels of co-authorship and number of articles per discipline for SSCI articles published in 2010.

Discipline	One-author	Two-author	Three-author	Four-author	Total articles
Business	25.7%	37.6%	27.3%	7.4%	8,553
Business, Finance	41.7%	41.1%	15.1%	1.8%	1,248
Economics	53.3%	34.9%	9.5%	1.7%	6,723
Education	46.5%	29.9%	12.9%	6.1%	5,201
Environmental Studies	35.1%	24.9%	17.0%	9.3%	4,945
Health Care	35.2%	22.7%	16.0%	11.0%	1,146
Health Policy	41.4%	22.6%	15.8%	8.3%	1,813
Information Science	69.9%	19.5%	7.3%	1.8%	2,336
International Relations	83.7%	13.4%	2.0%	0.6%	1,571
Law	78.8%	14.6%	3.8%	1.6%	2,909
Management	38.5%	39.9%	16.4%	4.1%	3,576
Nursing	47.5%	27.9%	13.1%	5.3%	1,669
Political Science	81.8%	14.0%	2.9%	1.1%	3,701
Psychology	32.4%	30.8%	18.6%	9.4%	17,125
Public Health	26.3%	22.6%	19.5%	12.4%	3,382
Rehabilitation	29.5%	30.2%	19.8%	11.9%	1,635
Social Sciences, Interdisciplinary	60.6%	23.9%	9.0%	3.5%	4,191
Sociology	64.8%	24.7%	7.7%	1.8%	2,587
Mean	49.6%	26.4%	13.0%	5.5%	4,128

4. Results

4.1. Variation in alphabetization over time (Question 1)

In answer to Question 1, although the percentage of co-authored articles varies substantially from discipline to discipline and in general exceeds the expected random value (ERV), there is little variation over time in the percentage of articles in alphabetical order (Tables 2 and 3). For each discipline the percentages of two-author, three-author and four-author alphabetized articles were calculated and expressed as a multiple of the ERVs (Table 2).

Table 2: Level of alphabetization expressed as a multiple of ERVs for SSCI articles published in 2010.

Discipline	Two-author	Three-author	Four-author	Mean
Business	1.36	2.82	6.24	3.47
Business, Finance	1.68	4.50	16.07	7.42
Economics	1.58	3.72	9.35	4.88
Education	1.06	1.14	1.20	1.13
Environmental Studies	1.06	1.20	1.68	1.31
Health Care	1.16	1.32	1.20	1.23
Health Policy	1.12	1.38	1.20	1.23
Information Science	1.04	1.38	2.88	1.77
International Relations	1.36	2.88	5.04	3.09
Law	1.38	2.22	2.64	2.08
Management	1.22	1.80	3.60	2.21
Nursing	1.00	0.90	1.20	1.03
Political Science	1.36	2.82	8.39	4.19
Psychology	1.02	1.08	1.20	1.10
Public Health	1.00	1.08	1.44	1.17
Rehabilitation	1.00	1.02	0.96	0.99
Social Sciences, Interdisciplinary	1.26	2.04	2.64	1.98
Sociology	1.10	1.56	2.16	1.61
Mean	1.21	1.94	3.84	2.33

Table 2 indicates that the percentage of alphabetized articles is 21% more than the ERV for two authors, 94% higher for three authors and 284% higher for four authors. However, there are very large variations between disciplines in the extent to which the level of alphabetization exceeds the ERV: Business, Finance exceeds ERV by 68% for two authors, 350% for three authors and 1,507% for four authors (although for only 21 four author articles), whereas Rehabilitation equals the ERV for two authors, exceeds the ERV by 2% for three authors and is 4% lower than the ERV for four authors. For four-author articles there is a substantial variation between disciplines in the extent to which the level of alphabetization exceeds the ERV: for five disciplines (Business, Finance; Economics; Political Science; Business; International Relations) this level exceeds the ERV by over 400% whereas for seven disciplines (Public Health; Education; Health Care; Health Policy; Nursing; Psychology; Rehabilitation) it is less than 45%. The Spearman correlation across disciplines between two authors and three authors is 0.95, between two authors and four authors is 0.79 and between three authors and four authors is 0.87 (all with $p < 0.01$). This suggests that the alphabetization practice is not dependent on the number of authors, but is affected to some extent by this; possibly highly co-authored articles may typically concern a different type of research from that of articles with two authors.

The mean percentages of alphabetized articles for articles published in 1995 are very similar to the mean percentages for articles published in 2010. This is demonstrated in Table 3, where the levels of alphabetization for 2010 are expressed as a multiple of the levels for 1995. A differences in proportions test was used to detect statistically significant changes in proportion between 1995 and 2010 (see Table 3).

Table 3: Levels of alphabetization for articles published in 2010 expressed as a multiple of the level for articles from 1995.

Discipline	Two-author	Three-author	Four-author	Mean
Business	0.97	1.01	0.87	0.95
Business, Finance	1.03	1.03	1.40	1.16
Economics	1.01	0.99	0.94	0.98
Education	0.99	0.86	0.61*	0.82
Environmental Studies	1.00	0.96	1.40	1.12
Health Care	1.01	1.00	1.04	1.02
Health Policy	1.11	1.18	1.06	1.12
Information Science	0.91	0.77	5.22+	2.30
International Relations	1.14*	1.53	0.63+	1.10
Law	1.05	1.51*	1.26+	1.28
Management	0.99	0.85*	0.52**	0.78
Nursing	0.95	0.97	0.75	0.89
Political Science	1.04	1.40*	1.71	1.38
Psychology	1.00	0.97	0.98	0.98
Public Health	1.02	1.03	1.58	1.21
Rehabilitation	0.99	0.92	0.87	0.93
Social Sciences, Interdisciplinary	1.03	1.21	1.07	1.11
Sociology	0.96	1.06	0.70	0.91
Mean	1.00	1.03	1.09	1.05

* changes significant at $p < 0.05$

** changes significant at $p < 0.01$

+ sample sizes too small to conduct a differences in proportions test

On average across disciplines and levels of authorship, the 2010 levels of alphabetization were only 5% higher than the 1995 levels (Table 3). However there were large variations between disciplines in the extent to which level of alphabetization changed between 1995 and 2010: (a) for two-author articles, this rose by 14% for International Relations but fell by 9% for Information Science (this fall was not significant), (b) for three-author articles, it rose by 51% for Law but fell by 15% for Management, and (c) for four-author articles, it rose by 422% for Information Science (this increase was based upon a small sample size and was not significant) but fell by 48% for Management and (d) the mean across levels of authorship rose by 130% for Information Science but fell by 22% for Management. The proportion test indicates that for only one discipline (Management) the change in the proportion between 1995 and 2010 was statistically significant for more than one of the levels of authorship. Spearman correlations between the 1995 and 2010 values were high at 0.90 for two-author, and 0.74 for four-author articles.

Finally, there was only a small variation with time in the means across disciplines. Using the data available on request, these mean values were calculated for each of the years 1995, 1998, 2001, 2004, 2007 and 2010 and ranged from 1.18 to 1.21 for two-author, from 1.74 to 1.91 for three-author, and from 3.45 to 4.03 for four-author articles.

4.2 First author skewing (Question 2)

In order to address Question 2, for each discipline the following frequencies were calculated for articles published in 2010: (a) the frequency of first author surnames of multi-authored articles that begin with A or B, C to E, F to H, I to K, L to M, N to R, S to T and U to Z; and (b) the frequency of author surnames of multi-authored articles that begin with A or B, C to E, F to H, I to K, L to M, N to R, S to T and U to Z, and (c) the ratios of frequencies in ‘(a)’ to the frequencies in ‘(b)’. These ratios are listed in Table 4 as percentages. For example, in Table 4 the ‘A – B’ column is calculated for multi-authored articles by:

$$\frac{100 \times (\text{the number of articles with } \textit{first} \text{ author surnames beginning with A or B})}{\text{the number of articles with } \textit{any} \text{ author surname beginning with A or B}}$$

The intervals were chosen to contain similar numbers of articles: across the disciplines the Mean percentages of all articles were respectively 13.3%, 14.7%, 15.2%, 8.6%, 13.7%, 12.8%, 12.4% and 9.4%. In order to demonstrate how skewing corresponds to level of alphabetization, the disciplines are presented in order of decreasing level of alphabetization, so that the discipline with the highest level of alphabetization for 2010 for two-author articles (Business, Finance) is first in the table. A difference in proportions test was used to detect statistically significant differences in the proportion of first author surnames beginning with A-B to the proportion beginning with U-Z. For example, for Business, Finance, this showed that the A-Z proportion of 0.741 was significantly different from the U-Z proportion of 0.176, with $p < 0.01$, taking into account the number of articles in the two bands used.

Table 4: For multi-authored articles in 2010 the number of first author surnames beginning with letters in eight bands expressed as a percentage of all multi-author surnames in the band (SSCI).

Discipline	A – B	C – E	F – H	I – K	L – M	N – R	S – T	U – Z
Business, Finance **	74.1%	60.8%	49.4%	42.7%	29.8%	25.3%	16.8%	17.6%
Economics **	65.6%	53.6%	44.8%	40.5%	32.8%	28.5%	22.5%	20.6%
Law **	50.0%	42.7%	40.9%	37.4%	35.8%	30.3%	30.8%	27.0%
Business **	52.8%	48.4%	43.1%	37.9%	33.0%	31.3%	26.6%	27.9%
International Relat. **	55.5%	48.2%	42.1%	36.9%	36.3%	33.5%	27.4%	28.0%
Political Science	58.4%	51.0%	44.6%	38.6%	38.0%	34.1%	31.7%	28.0%
Social Sciences, Inter. **	33.7%	32.0%	31.2%	30.4%	27.5%	27.5%	24.7%	25.1%
Management **	43.4%	40.8%	38.9%	35.9%	34.3%	32.9%	31.1%	31.6%
Health Care **	21.7%	21.8%	20.2%	22.8%	21.6%	19.9%	20.1%	18.1%
Health Policy	22.5%	23.0%	22.3%	24.6%	22.6%	21.2%	21.3%	20.3%
Sociology **	43.7%	40.7%	38.4%	38.3%	34.4%	36.3%	35.9%	35.5%
Education *	32.6%	32.1%	30.7%	31.2%	31.3%	29.2%	27.9%	29.8%
Environmental Studies	24.0%	23.2%	22.7%	24.3%	22.4%	23.0%	22.8%	23.6%
Information Science	37.2%	33.9%	29.0%	35.0%	36.8%	30.5%	29.2%	34.0%
Psychology	26.5%	27.2%	26.5%	27.6%	26.3%	26.0%	26.0%	27.4%
Nursing	26.7%	26.7%	26.8%	28.0%	26.5%	26.2%	25.7%	24.5%
Public Health	21.3%	20.9%	20.5%	22.2%	20.4%	21.6%	20.9%	21.8%
Rehabilitation	24.7%	23.0%	25.0%	26.3%	25.2%	21.7%	23.7%	24.5%
Mean	39.7%	36.1%	33.2%	32.3%	29.7%	27.7%	25.8%	25.9%

* significant at $p < 0.05$

** significant at $p < 0.01$

Table 4 indicates a strong association between the skewing of first-authorship of multi-authored articles towards author surnames beginning higher in the alphabet and the level of alphabetization. For the most highly alphabetized discipline (Business, Finance) the percentage of multi-authored first authorship for surnames beginning with A-B was more than 320% higher than the percentage for surnames beginning with U-Z (74.1/17.6) whereas for the three disciplines with the lowest level of alphabetization was on average only 3%

higher. Spearman correlations between the percentages of first authors with surnames beginning with A-B and the levels of alphabetization for 2010 (Table 2) were calculated; the correlations were 0.79 for two-author alphabetization, 0.86, for three-author, and 0.86 for four-author alphabetization (all with $p < 0.01$).

First author skewing was also addressed for twenty further disciplines. Of the 25 disciplines identified by Waltman (2012) as having the highest level of alphabetization, five were investigated in the current study (Business, Finance; Economics; International Relations; Law and Political Science). For the other twenty disciplines total percentages of first authorship corresponding to those in Table 4 were calculated and the findings presented in Table 5. In Table 5, in order to demonstrate how skewing is associated with level of alphabetization, the disciplines are presented in decreasing level of alphabetization (called ‘% alphabetical pub.’ by Waltman) of the data for 2010 for two-author articles. Each discipline was taken from the appropriate part of WoS.

Table 5: For multi-authored articles in 2010 the number of first author surnames beginning with letters in eight bands expressed as a percentage of all multi-author surnames in the band (WoS).

Discipline	A – B	C – E	F – H	I - K	L - M	N – R	S - T	U - Z
Mathematics **	63.7%	51.6%	43.5%	40.6%	34.7%	26.1%	22.0%	27.4%
Philosophy	48.9%	46.0%	42.5%	47.9%	40.4%	32.8%	34.8%	40.2%
Social Sciences, Mathem. **	52.9%	53.7%	44.9%	41.4%	32.6%	29.0%	24.0%	22.3%
Physics, Particles **	13.5%	8.9%	7.5%	6.6%	5.7%	4.1%	3.9%	5.2%
Mathematics, Applied **	63.2%	50.7%	42.6%	39.7%	34.7%	26.6%	23.2%	29.3%
History **	46.0%	46.4%	40.3%	45.3%	33.9%	36.6%	39.7%	24.7%
Statistics **	44.7%	40.7%	38.2%	36.5%	31.7%	29.9%	25.1%	29.9%
Area Studies **	51.8%	50.0%	50.5%	42.3%	40.0%	41.2%	32.3%	30.0%
Industrial Relations **	52.0%	44.5%	39.6%	33.1%	29.9%	29.9%	28.5%	31.9%
Physics, Mathematical **	47.2%	39.1%	35.5%	32.4%	30.0%	26.9%	24.7%	27.9%
Planning **	48.8%	44.1%	41.5%	35.2%	34.9%	29.8%	32.8%	30.2%
Humanities, Multidiscip. **	42.0%	53.5%	39.3%	47.1%	41.0%	33.0%	36.8%	40.4%
Computer Science Theory **	47.5%	37.8%	33.2%	31.3%	27.2%	23.8%	24.9%	27.0%
Language & Linguistics **	49.0%	43.4%	41.6%	36.6%	36.8%	37.7%	34.1%	35.9%
Public Administration **	45.7%	43.7%	40.2%	38.8%	38.9%	34.8%	27.7%	31.5%
History & Philosophy	43.0%	36.6%	41.6%	50.5%	38.6%	34.1%	34.3%	36.6%
Urban Studies **	47.7%	38.0%	38.1%	32.0%	40.3%	32.7%	31.5%	29.4%
Operations Research	47.8%	38.8%	36.0%	35.6%	32.1%	29.2%	28.4%	31.5%
Demography	33.3%	41.4%	35.6%	34.7%	35.2%	33.5%	31.6%	29.7%
Mathematics, Interdisciplinary **	42.7%	37.8%	34.8%	37.3%	33.1%	29.6%	27.6%	29.8%
Mean	46.6%	42.3%	38.3%	37.2%	33.6%	30.1%	28.4%	29.5%

** significant at $p < 0.01$

Table 5 indicates a strong association between the first-authorship skewing of multi-authored articles towards author surnames beginning higher in the alphabet and the level of alphabetization. The correlation between the percentage of first authors to authors with surnames beginning with A-B and Waltman’s % alphabetical pub. is 0.48 ($p < 0.05$). It seems natural for the correlation between skewing and alphabetization to be lower for these disciplines than for the disciplines in Table 4, as the disciplines in Table 5 are not a complete or random selection but disciplines selected for alphabetization.

5. Discussion

The extent to which the level of alphabetization exceeds the ERV varies considerably from discipline to discipline and is more marked for higher levels of authorship. For 2010

alphabetical order for Rehabilitation is 4% lower than the ERV for four authors and equals the ERV for two authors, whereas for Business, Finance it exceeds ERV by 1,507% for four authors and 68% for two authors. This much larger divergence from the ERV for a larger number of authors is mirrored by the behaviour of the mean across disciplines; the mean alphabetization exceeds the ERV by 284% for four authors and only 21% for two authors. These findings indicate that in some disciplines the lists of co-authors are deliberately put into alphabetical order; this practice is particularly evident for four-author articles, which are unlikely to be in alphabetical order by chance.

The mean alphabetization across disciplines and level of authorship only rose by 5% between 1995 and 2010. As discussed above, in Economics Frandsen and Nicolaisen (2010) found a slight increase in the percentage of alphabetized articles in 2007 compared to 1988; we found that for Economics the levels of alphabetization in 1995 and 2010 were almost identical (the 2010 values for two-author, three-author, and four-author were respectively 101%, 99% and 94% of the 1995 values. In Information Science Frandsen and Nicolaisen (2010) found a sharp decrease in the percentage of alphabetized articles from 1988 to 2007; we found that for Information Science the 2010 values for two-author articles and three-author articles were lower than the 1995 values (91% and 77% respectively) whereas for four-author articles the value for 2010 was much higher than for 1995 (522%). This apparently strange contrast may reflect an influx of highly co-authored research within the discipline (e.g., computer science) and hence a different dynamic for different levels of co-authorship.

Of the 25 categories identified by Waltman (2012) as having the highest level of alphabetization, five were investigated in this current study (Business, Finance; Economics; International Relations; Law and Political Science). These five categories together with Business are the six categories in this current study with the highest level of two-author articles in alphabetical order. Waltman's (2012) finding of a limited change in alphabetization over time in four categories (Business, Finance; Economics; Mathematics; Physics, Particles & Fields) was extended in the current study; we found over 1995 to 2010 little change in the level of alphabetization not only for Business, Finance and Economics but for social science disciplines in general.

As described in the 'Related literature and research questions' section, Einav and Yariv (2006) found that tenured faculty in Economics had, in general, surnames higher in the alphabet than untenured faculty. The data in Table 4 confirms that alphabetization is a likely explanation: In Economics for 2010 first authors with surnames beginning with A or B are 2.35 times more often the first author of two-author articles than authors with surnames beginning with the letters U, V, W, X, Y or Z and the allocation of tenure could be influenced by the number of first-author articles. Einav and Yariv found that 88% of the multi-authored articles published in five prominent economics journals during 1980-2002 authors were alphabetized; we found that 75.6% of the two-author articles published in 2010 were alphabetized.

The findings on the percentages of alphabetized articles need to be interpreted with care. For instance, although in 2010 83.7% of the two-author articles in Business, Finance had authors' surnames in alphabetical order it is a justifiable, but unproven, assumption that the amount above the ERV indicates the extent to which articles were deliberately put in alphabetical order. The justification for the assumption is that five of the eighteen disciplines have a percentage very close to the ERV and it does not seem plausible that factors other than deliberately allocating alphabetical order would have such a dramatic effect on the percentage for Business, Finance and very little effect on the five disciplines that are very close to the ERV. However, it is possible that gender and nationality factors may impact on author-order and have a second-level effect on alphabetical order. For instance, common Chinese

surnames often start with letters lower in the alphabet (according to a translation², the 10 most frequent surnames in China in 2007 are Wáng, Zhāng, Liú, Chén, Yáng, Huáng, Zhào, Wú, Zhōu and Xú) and this might have resulted in alphabetical ordering of authors when the Chinese author was a subsidiary author to a non-Chinese author.

6. Conclusions

This research finds for social science clear-cut answers to the research questions. In response to question 1, there are substantial differences between disciplines in the extent to which alphabetization varies over time, with Management and Education tending to decrease and International Relations, Law, and Political Science tending to increase alphabetization. A limitation of this longitudinal investigation is that some differences may be due to changes in the set of journals covered by Web of Science through the addition or removal of journals. However, on average, the variation from year to year was relatively minor and so it seems that previous studies demonstrating the dangers of alphabetization have had little impact in practice.

Question 2 asks: Is there any relationship between a discipline having a high percentage of alphabetized articles and the skewing of the surnames of first authors towards the beginning of the alphabet? The anticipated relationship between a high level of alphabetization and skewing of first authorship towards the beginning of the alphabet was confirmed by the very high Spearman correlation of 0.79 with two-author alphabetization.

This study has ramifications for attributing credit to co-authors: When the level of alphabetization substantially exceeds the ERV, the assumption that author contribution reflects author order is particularly questionable when author surnames are listed alphabetically. It also has ramifications when assigning credit to authors: Authors with surnames closer to the beginning of the alphabet are more likely to be first named author than those near the end of the alphabet, and this affect is particularly marked for disciplines with a high percentage of alphabetized articles.

Given the impact of alphabetization on careers and the fact that it is slightly increasing rather than decreasing, those involved in research evaluation, promotions and appointments for disciplines like Economics and Business, Finance should take particular care not to assume that author order reflects contribution. Moreover, since alphabetization seems to be almost universally present to some extent in the social sciences, it seems likely that it is having at least a small impact on careers in many, if not most social sciences. Hence, in order to be able to more effectively allocate research credit, it seems that, in cases where the journal does not specify the relative contribution of the co-authors or where this information is not made readily available in CVs and employed in any research metrics used, alphabetization should be actively discouraged. For example a clear statement in a journal's guidance for authors that this practice is unacceptable may help to make a difference.

7. Acknowledgements

This research is supported by the Economic and Social Research Council [grant reference: RES-000-22-4415]. We thank Gertrude Levitt and the reviewers for very helpful feedback and comments on the drafts.

² http://en.wikipedia.org/wiki/List_of_common_Chinese_surnames

8. References

- Baerlocher, M.O., Newton, M., Gautam, T., Tomlinson, G. & Detsky, A.S. (2007). The meaning of author order in medical research. *Journal of Investigative Medicine*, 55(4) 174-180.
- Birnholtz, J.P. (2007). When do researchers collaborate? Toward a model of collaboration propensity. *Journal of the American Society for Information Science and Technology*, 58(14) 2226-2239.
- 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. *British Medical Journal*, 323(7327) 1460-1461.
- Costas, R. & Bordons, M. (2011). Do age and professional rank influence the order of authorship in scientific publications? Some evidence from a micro-level perspective. *Scientometrics*, 88(1) 145-161.
- Davenport, E. & Cronin, B. (2001). Who dunnit? Metatags and hyperauthorship. *Journal of the American Society for Information Science and Technology*, 52(9) 770-773.
- Einav, L. & Yariv, L. (2006). What's in a Surname? The Effects of Surname Initials on Academic Success. *The Journal of Economic Perspectives*, 20(1) 175-187.
- Engers, M., Gans, J.S., Grant, S. & King, S.P. (1999). First-author conditions. *Journal of Political Economy*, 107(4) 859-883
- Frandsen, T.F. & Nicolaisen, J. (2010). What is in a name? Credit assignment practices in different disciplines. *Journal of Informetrics* 4(4) 608-617.
- Good I.J. (1989). Order of authorship. *American Statistician*, 42(4) 288-289.
- Joseph, K., Laband, D.N. & Patil, V. (2005). Author order and research quality. *Southern Economic Journal*, 71(3) 545-555
- Hagen, N.T. (2010). Harmonic publication and citation counting: sharing authorship credit equitably – not equally, geometrically or arithmetically. *Scientometrics*, 84(3) 785-793.
- Hilmer, C.E. & Hilmer, M.J. (2005). How do journal quality, co-authorship, and author order affect agricultural economists' salaries? *American Journal of Agricultural Economics*, 87(2) 509-523.
- Hodge, S.E. & Greenberg, D.A. (1981). Publication credit. *Science*, 213(4511) 950.
- Laband, D. & Tollison, R. (2006). Alphabetized coauthorship. *Applied Economics*, 8(14) 1649-1653.
- Laband, D.N. (2002). Contribution, attribution and the allocation of intellectual property rights: Economics versus agricultural economics. *Labour economics*, 9: 125-131.
- Marusic, A., Bosnjak, L. & Jeroncic, A. (2011). A Systematic Review of Research on the Meaning, Ethics and Practices of Authorship across Scholarly Disciplines. *Plos One*, 6(9) article e23477.
- Rudd, E. (1977). Effect of alphabetical order of author listing on careers of scientists. *Social Studies of Science*, 7(2) 268-269.
- Savitz, D.A. (1999). What can we infer from author order in epidemiology? *American Journal of Epidemiology*, 149(5) 401-403.
- Tol, R.S.J. (2011). Credit where credit's due: accounting for co-authorship in citation counts. *Scientometrics*, 89(1) 291-299.
- van Praag, C.M. & van Praag, B.M.S. (2008). The benefits of being economics professor A (rather than Z). *Economica*, 75: 782-796.
- Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, 6(4), 700–711.