

Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers¹

Mike Thelwall and Nabeil Maflahi

Statistical Cybermetrics Research Group, University of Wolverhampton.

International collaboration tends to result in more highly cited research, and, partly as a result of this, many research funding schemes are specifically international in scope. Nevertheless, it is not clear whether this citation advantage is the result of higher quality research or is due to other factors, such as a larger audience for the publications. To test whether the apparent advantage of internationally collaborative research may be due to additional interest in articles from the countries of the authors, this article assesses the extent to which the national affiliations of the authors of articles affect the national affiliations of their Mendeley readers. Based upon English-language Web of Science articles in ten fields from science, medicine, social science, and the humanities, the results of statistical models comparing author and reader affiliations suggest that, in most fields, Mendeley users are disproportionately readers of articles authored from within their own country. In addition, there are several cases where Mendeley users from certain countries tend to ignore articles from specific other countries, although it is not clear whether this reflects national biases or different national specialisms within a field. In conclusion, research funders should not incentivise international collaboration on the basis that it is, in general, higher quality because its higher impact may be primarily due to its larger audience. Moreover, authors should guard against national biases in their reading in order to select only the best and most relevant publications to inform their research.

Introduction

In theory, many areas of the natural and logical sciences, medicine, and social sciences are inherently international, with researchers collectively attempting to understand the world and building an edifice of knowledge through their published works. International collaboration has been found to be particularly successful in terms of increased citations to publications, which may have prompted its promotion by policy-makers and funders (e.g., Lasthiotakis, Sigurdson, & Sá, 2013). Nevertheless, it is not clear that the increased number of citations attracted by internationally collaborative research is due to its higher quality rather than to other factors, such as the nature of the problem addressed, or the amount of funding received. Hence it is important to investigate all factors that may explain the higher impact of internationally collaborative research. One such factor is the visibility of co-authored publications: in theory, internationally co-authored research may have an advantage because it may be noticed more in each of the countries of the authors. This hypothesis is investigated here.

Although it is not possible to determine the nationality or country affiliation of all readers of academic articles, this article exploits a relatively new data source, Mendeley readers, to give new insights into the relationship between the nationalities of the users and authors of articles. The social reference sharing website Mendeley is free and allows people who sign up to list articles in their profiles that they have read, or intend to read, although no checking is done to verify that users are genuinely interested in their listed articles. Some users also register a country location in Mendeley and hence it is possible to cross-reference the countries of the readers of articles in Mendeley with the countries of the authors of those articles. This data is analysed for ten subjects,

¹ This is a preprint of an article published in the Journal of the Association for Information Science and Technology © copyright 2015 John Wiley & Sons, Inc. DOI: 10.1002/asi.23252

chosen to be disparate but with large Mendeley readerships, in order to check whether these readers tend to come from the same countries as the authors of the articles.

Background

International collaboration

International collaboration is the co-authoring of publications with author affiliations coming from at least two different countries. Although collaboration can take other forms that do not lead to co-authorship (Cronin, Shaw, & Barre, 2004; Katz, & Martin, 1997) and co-authors do not always contribute to a paper (Baerlocher, Newton, Gautam, Tomlinson, & Detsky, 2007), co-authorship seems to be a reasonable general approximation to substantial collaboration. International collaboration is common in science although the extent to which it occurs varies greatly between fields and countries (e.g., Luukkonen, Persson, & Sivertsen, 1992). Geography is a factor in collaboration, with academics tending to collaborate at a much higher rate with those near them, both within their own country (Katz, 1994) and abroad (Luukkonen, Persson, & Sivertsen, 1992).

Internationally co-authored papers tend to be more highly cited than other papers. For example, more countries represented within the authorship team associates with more citations (Bote, Vicente, Olmeda-Gómez, & Moya-Anegón, 2013; Didegah & Thelwall, 2013), although the most highly cited articles within a field are not necessarily highly international (Persson, 2010). At the level of individual sciences and countries, international collaboration tends to associate with more citations, although there are a few exceptions (Glänzel, 2001) and some countries benefit more from collaboration than do others (Lancho-Barrantes, Guerrero-Bote, & de Moya-Anegón, 2013). The reason for the additional citations for internationally collaborative research is unclear, however. For example, it may be that the most highly funded projects tend to be international in scope (e.g., Luukkonen, Persson, & Sivertsen, 1992) or that the best authors prefer to collaborate internationally with the best authors in the rest of the world. A confounding factor is the number of authors because in many fields more highly co-authored articles tend to be more highly cited (Francescheta & Costantini, 2010; Wuchty, Jones, & Uzzi, 2007). The reason for the citation advantage of collaboration is not fully understood. It could be that larger teams are needed to deal with more difficult problems, the solutions of which then get highly cited, or that additional authors generate additional expertise that leads to better papers (e.g., Katz & Hicks, 1997; Price, 1963; Gibbons et al., 1994). Although it is known that high impact authors and institutions are highly collaborative (Gazni, Sugimoto, & Didegah, 2012), it is not clear whether the collaboration causes the impact or vice versa. Conversely, more authors may generate more self-citations, although this probably does not account for all of the increased impact of collaborative papers (Glänzel, & Thijs, 2004; van Raan, 1998).

International collaboration seems to be more beneficial than national collaboration for citations (e.g., Persson, Glänzel, & Danell, 2004), although another paper using different methods suggests the opposite, at least for biochemistry: that the key factor is the number of authors rather than the number of countries (Sud & Thelwall, submitted). Hence, whilst it is agreed that increasing the number of countries seems to be beneficial, partly because it increases the number of authors, it is not clear whether independently increasing the number of countries is also beneficial. These studies used different methods and both analysed multiple fields and the difference in conclusions could be associated with the simplifying statistical assumptions made in either paper or could reflect changes over time. In any case, other types of evidence are still needed.

One reason why it seems plausible that international research may be more highly cited than national research is that, irrespective of the quality of an article, each author from a different country tends to increase the visibility of an article within their country. Although there is little direct evidence of this (e.g., in terms of international differences in article downloads), there is indirect evidence in that scientists tend to collaborate with others in the same country (Hoekman, Frenken, & Tijssen, 2010) and tend to cite works authored within their own country (Lancho-Barrantes, Bote,

Vicente, Rodríguez, & de Moya Anegón, 2012), including in patents (Jaffe, & Trajtenberg, 1999). Hence it seems reasonable to believe that collaborative research could attract particularly large *audiences* within the countries of the authors, although it is not clear whether the magnitude of this effect is substantial enough to be worth considering.

There are many reasons why authors may tend to cite articles from their own country more and these reasons also suggest that papers may have a larger audience in their own country. Intuitively, although science is international it has national components that may make research more visible within the country of the authors. These include press coverage of national scientific discoveries, and disproportionate press coverage of a country's own scientists. Moreover, the tendency for countries to specialise to some extent in specific fields (e.g., Schwarz, 1999) or in specialisms within fields means that authors may tend to cite others from the same country just because they are more likely to work in the same specialism or field. Another reason why authors tend to cite others from the same country is that they seem to cite former collaborators more (Wallace, Larivière, & Gingras, 2012), and these are more likely to be from their own country than from other countries. Finally, a tendency for national self-citation (i.e., articles citing other articles with authors from the same country) could, in theory, also occur if some countries tended to publish in high impact journals and cite other high impact journals, whereas other countries tended to publish in low impact journals and cite low impact journals. This does not seem to happen, however, since even articles in the lowest strata of web of science journals do not cite articles in the top strata of journals less than average (Didegah, Thelwall, & Gazni, 2012).

Mendeley

The social reference sharing site Mendeley allows registered users to share their references with other users (Gunn, 2013; Henning & Reichelt, 2008; Zaugg, West, Tateishi, & Randall, 2011). A by-product of this is that, for any paper, Mendeley records the number of users that have listed it, describing them as *readers*, whether or not they actually read it. Presumably, listing an article in Mendeley tends to reflect that an article has been read or will be read in the future, although there is no evidence that this assumption is true. The reader counts for an article can thus give useful information about its uptake. The number of readers of an article has been found to correlate with the number of citations to it for a number of fields and journals (Bar-Ilan, 2012; Li, Thelwall, & Giustini, 2012; Li & Thelwall, 2012; Mohammadi & Thelwall, in press; Thelwall, Haustein, Larivière, & Sugimoto, 2013), confirming that readership counts are probably related to scholarly activities and also that they can potentially be used as early indicators of likely future citation counts. This justifies the use of Mendeley reader counts as an *altmetric* (Priem, & Hemminger, 2010; Priem, Taraborelli, Groth, & Neylon, 2011; Torres, Cabezas, & Jiménez, 2013), a social web indicator of an aspect of the value of academic articles. Moreover, at least in the case of one journal, Mendeley readers correlate more highly with downloads than with citations, suggesting that Mendeley readers may reflect wider usage of articles than that of just academics (Schlögl, Gorraiz, Gumpenberger, Jack, & Kraker, 2013). As background to this, downloads correlate with citations in many areas of science (Schlögl, & Gorraiz, 2011; Moed, 2005; Kurtz, & Bollen, 2010), especially when relatively few downloads come from undergraduates (Bollen, & Sompel, 2008). A final justification for the use of Mendeley is that it seems to give higher values than other altmetrics for Web of Science (WoS) articles (Zahedi, Costas, & Wouters, 2013a).

Mendeley readers are not the same as citers because they are both a narrower and a broader group. Mendeley readers are a narrower group in the sense that, unlike the majority of academic authors, they are people that use Mendeley. Presumably this makes them younger, more sociable and more technologically-oriented than average for researchers. Mendeley users are also a broader group in the sense that many may not be authors of academic papers but may be students, professionals or others. Hence Mendeley readership may reflect broader impacts of academic research, albeit within a biased subset of the scientific audience, perhaps tending to reflect educational rather than scholarly uses.

An important additional source of bias within Mendeley is geographic. Social websites seem to have strong patterns of use by nation. For example, whilst some social websites are dominant within one country (e.g., VK.com gets 67% of its visitors from Russia and is the most popular social website in Russia, according to <http://www.alexa.com/siteinfo/vk.com>) they may be unused in others. Hence, Mendeley is also likely to have differing levels of uptake across the globe. In support of this, Alexa.com suggests that its largest ten users are the USA (16.1%), India (13.2%), Belgium (9.9%), Germany (6.2%), UK (5.9%), Japan (4.6%), France (2.8%), Brazil (2.8%), Australia (2.3%), and Spain (2.1%) (<http://www.alexa.com/siteinfo/mendeley.com>, as of 14 December 2013). Moreover, since science is organised by discipline, Mendeley is also likely to have differing levels of uptake between fields. For example, a sample of 74,519 WoS articles from 2011-2012 with at least 1 Mendeley reader (by early 2013) found that social and behavioural sciences articles had five times as many Mendeley readers as citations per paper whereas medical and life sciences articles had only 1.6 times as many Mendeley readers as citations per paper (Zahedi, Costas, & Wouters, 2013b).

Research questions

The objective of this paper is to assess the extent to which Mendeley readership associates with the authorship of articles in order to shed more light on the relationship between the authorship and readership of a paper, and on the value of international collaboration. Mendeley is used as a free and universal source of evidence about the readership of articles. Publishers do not provide this information, although scholarly download information is sometimes available to insiders within restricted contexts (e.g., Bollen, & Sompel, 2008; Duin, King, & van den Besselaar, 2012), and other sources of free article readership information, such as ResearchGate and other reference sharing sites, seem to be less comprehensive than Mendeley. The following research questions drive the investigation.

1. Do articles attract more Mendeley readers from a country if at least some of the article authors come from that country?
2. Do articles from specific countries tend to attract disproportionately many (or few) Mendeley readers from any specific other countries?

Data and methods

The research questions were addressed for a diverse set of fields to ensure that the findings would be widely relevant. The fields were chosen for having relatively extensive Mendeley use compared to similar fields. The amount of Mendeley use was judged by the size of the subject categories in the Mendeley website and so, in practice, the fields were chosen from those listed in Mendeley as having large numbers of articles. This is a heuristic, however, because it depends upon Mendeley's categories – for example, a category could halve in size if split into two separate named groups. Fields were matched as far as possible with WoS categories in order to gather complete lists of articles from a coherent collection of relevant journals. WoS was chosen in preference to Scopus but either would have been acceptable. Although WoS is known to have international biases, such as language biases (Van Leeuwen, Moed, Tijssen, et al., 2011), and Scopus appears to have reasonable international coverage, albeit not perfect (de Moya-Anegón, Chinchilla-Rodríguez, Vargas-Quesada, et al., 2007), the analysis method chosen should minimise the effect of these, and country-level indicators are not affected by the choice of database (Archambault, Campbell, Gingras, & Larivière, 2009) so the impact of this decision seems unlikely to be large. Nevertheless, it is possible that the results would differ for equivalent Scopus data in fields in which Scopus had substantially greater coverage, such as Nursing, for example. Presumably, a Scopus dataset with wider coverage outside the core set of international journals would increase the chance of finding a positive result from the statistical tests used here, and so the use of WoS instead of Scopus should not cause spurious evidence of authors tending to read articles from their own country; on the contrary, it might hide such a tendency if it is mostly evident in the national literature. Similar categories were avoided in

order to give a spread of subjects. Categories with obvious direct connections to national differences (e.g., law, economics, agriculture) were avoided, although the remaining more applied categories could be connected to national organisational structures (e.g., education, nursing, library science) or national geographic characteristics (e.g., ecology) to some extent. The final list was: Biochemistry; Information Science & Library Science (ISLS); Ecology; Education, Educational Research (Education); Computer Science, Artificial Intelligence (AI); Nursing; Oncology; Engineering, Mechanical (MechEng), Neuroscience; and Philosophy.

For each selected WoS category, all documents of type article (only) that were written in English in 2011 were downloaded. The year 2011 was selected to give some time for articles to attract readers. A longer period of time was not considered necessary because social web mentions seem to be faster to accrue than citations. Only English articles were analysed because English is currently the international language of science (e.g., Liang, Rousseau, & Zhong, 2013) and allowing multiple languages would add a confounding factor to the analysis. Articles with dual subject classifications were excluded to avoid contaminating the data with multidisciplinary articles.

The country affiliation of the authors was automatically extracted from the WoS affiliation field (C1). Cases where there were different numbers of authors in the affiliation field and in the author field were discarded. Only the first affiliation of each author was considered. The country was determined from the last segment of each affiliation address. US states were removed from affiliations and all UK countries were combined into a single UK group. For each article and each nation, the national authorship proportion was calculated as the ratio of national to all authors. For example, if 3 out of 7 authors were from the UK then the UK authorship proportion would be 3/7.

Each article was searched for in Mendeley using the full article title as a quoted search, together with the publication year and the last name of the first author. This seemed to give no false matches but some multiple matches for duplicate records for the same article. When duplicate records were found, only the record with the largest number of readers was kept.

For each article, the number of readers from each country was extracted as far as possible. Reader national location information is only partially available for two reasons. First, most users do not record their country and so only some readership country location information is available. Second, Mendeley records only the top 3 nationalities for each article. In practice, the latter was rarely a problem because few articles had Mendeley readers from more than 3 declared nationalities due to the small numbers involved and the high proportion of readers not declaring a country. The number of readers from each country was calculated by multiplying the total number of readers for an article by the reported percentage of readers from that country, rounding to the nearest whole number.

For each of the selected fields, the ten countries with the most first authors in the articles from 2011 were identified. For each of these ten countries, a separate count data model was built to regress the number of Mendeley readers from that country against the proportion of authors from each of the top 10 countries. Three additional variables were also added to the model to account for different types of collaboration known to affect the citation impact of research: the number of authors, the number of institutions and the number of countries represented. Hence, each model had 13 predictor variables, all calculated from each article's authors, and one independent variable, the Mendeley readership from a particular country. The inclusion of all ten countries in each regression model minimised the risk that there would be false positive answers to the first research question due to readers focusing on articles from a small number of countries rather than just from their own country.

There are many different count data regression models that could be used to model the data. The natural choice is the Zero-Inflated Negative Binomial (ZINB) (Hilbe, 2011) because the data contains many zeros in the independent variable (readers from a country) and some of these zeros are artificial in the sense that they correspond to articles not found by the Mendeley search process due to typos or strange characters in article titles. Nevertheless, the number of artificial zeros from this source may be small enough to ignore. Additional zeros are also present for readers who did not

declare a country location but this is not necessarily a process that needs to be modelled separately since, assuming that there is little pattern in the people that do not declare a country, it would uniformly reduce all the proportions and could be modelled without zero inflation. Hence, although ZINB regressions were used as the main tests in all cases, secondary tests with standard Negative Binomial (NB) and Zero Inflated Poisson (ZIP) models were also calculated and compared with the ZINB results to check that the reported results were not influenced by the use of ZINB in comparison to alternative plausible distributions. For the zero inflation component of ZIP and ZINB only the number of authors was used as a predictor since the zero generation seems to be an almost random process and using all the variables would therefore needlessly reduce the power of the main regression component of the model. A second, simpler ZINB model was also fitted as a check, with the zero inflation component containing only a constant and the department variable moved – the latter because it strongly correlated with the institutions variable and hence may stop the model fitting from converging in some cases. In situations where the main model could not be fit, the result was substituted with the next most appropriate model. This seemed reasonable because when different models fitted the same data, their results were almost always identical and in the situations where the results differed, these differences were only minor (e.g., switching from just above the 5% confidence limit to just below).

Some of the Mendeley readers of articles may also be authors of the articles and if there are many of these *self-readers* then they could bias the results in favour of readers tending to come from the same country as authors. It is difficult to filter out self-readers because Mendeley does not list an article's readers and it is impractical to find each author's Mendeley profile, if any, and then check if they list their own articles. Instead, a test was performed to see if potential authors were over-represented in the readers of articles from the same country. For each article, Mendeley lists the proportion of readers with the top 3 types of academic status, and three of these seem very unlikely to be authors: *Librarian* (except perhaps in IS&LS), *Student (Master)* and *Student (Batchelor)*, although the others could be (e.g., *Ph.D. Student*, *Assistant Professor*). This information seems to be present in all cases, in contrast to reader country information. Hence, for each article with at least one reader, the number of readers that were librarians, undergraduates or masters students was calculated as well as the total readers of all the other types. Since Mendeley does not provide the raw data in each case but only the total number of readers and the percentage of the most common three types of reader, the totals were calculated by multiplying the percentages by the total number of readers (i.e., the same process as for author nationalities). Some information will be lost for all articles with more than three different academic statuses of reader, but this seems unlikely to mask any systematic sources of bias. A difference in proportions test was applied to see if the proportion of the non-author category was significantly different for papers with at least one author from the same country in comparison to the proportion for all other papers. In cases where the difference was statistically significant, the difference between the proportions was calculated and reported as the *author bias*. For example, an author bias of 6% for a country indicates that the proportion of readers in the potential author categories was 6% larger for articles authored by someone from the country in comparison to other articles.

Results

Tables 1-10 report the results of the tests for each field². Each table summarises the results of all regression models for one of the selected fields, showing the statistically significant national relationships. The results would be almost identical if the NB, ZIP or simple data ZINB models had been used, with the exceptions that NB or ZIP models were possible in some cases for which ZINB models did not converge, presumably due to too little data in most cases, but perhaps due to poor

² Statistical tests available at http://figshare.com/articles/Are_scholarly_articles_disproportionately_read_in_their_own_country_An_analysis_of_Mendeley_readers/902197

model fits in others. In all tables, black indicates a statistically significant positive relationship and grey indicates a statistically significant negative relationship ($p < 0.05$ in both cases). White squares indicate the lack of a statistically significant relationship. Each column represents a single model for the number of readers from a specific country and each row represents the proportion of authors for articles from the named country. For example, the black cell in the USA row and UK (M5) column of the AI table indicates that Model 5, a ZINB regression model for the number of UK readers of articles, found that the proportion of article authors from the USA was a statistically significant and positive predictor of the number of UK readers. In other words, after taking the other factors in Model 5 into account, the higher the proportion of US authors of AI papers, the higher the average number of UK readers they were likely to have.

Table 1. Biochemistry. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

n=6745		Location of readers										Papers	Author bias
Model	M1	M2	M3^	M4	M5^	M6^	M7	M8	M9	M10^			
		USA	UK	Germany	China	Japan	France	Canada	Italy	Spain	India		
Authorship (proportions)	USA	Black	Black	Black		Black						3112	2%
	UK	Black	Black									647	-3%
	Germany	Black		Black				Black				641	
	China	Grey	Grey	Grey	Black							634	
	Japan					Black						477	
	France						Black	Black	Black			370	-7%
	Canada		Black				Grey	Black				330	
	Italy	Grey							Black			198	
	Spain			Black						Black		198	4%
	India	Grey	Grey								Black	177	-8%
	Readers	5302	2233	1555	478	829	812	607	308	433	408		

^= simple ZINB replacing standard ZINB

Table 2. Neurosciences. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

		Location of readers											
		M1	M2	M3^	M4	M5	M6	M7	M8	M9	M10^	Papers	Author bias
		USA	Germany	Japan	UK	Canada	China	Italy	France	Netherlands	Spain		
Authorship (proportions)	USA	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	2643	2%
	Germany	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	640	-1%
	Japan	Black	Black	Black	Grey	Black	Black	Black	Black	Black	Black	498	
	UK	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	484	
	Canada	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	409	-4%
	China	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	403	-6%
	Italy	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	299	
	France	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	279	
	Netherl.	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	193	-2%
	Spain	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	155	
Readers		9459	3899	1462	3453	626	372	450	733	812	286		

^= simple ZINB replacing standard ZINB

Table 3. Oncology. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

		Location of readers											
		M1	M2	M3	M4	M5	M6^	M7	M8	M9	M10	Papers	Author bias
		USA	China	Japan	UK	Germany	Italy	France	Canada	Netherlands	S. Korea		
Authorship (proportions)	USA	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	1900	3%
	China	Grey	Black	Black	Black	Black	Black	Black	Black	Black	Black	549	
	Japan	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	443	-4%
	UK	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	399	
	Germany	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	372	
	Italy	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	290	5%
	France	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	272	
	Canada	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	250	
	Netherl.	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	184	
	S. Korea	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	179	-22%
Readers		1684	233	550	1122	424	163	226	365	191	72		

^= simple ZINB replacing standard ZINB

Table 4. Ecology. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

n=2186		Location of readers										Papers	Author bias	
Model	M1	M2	M3	M4	M5	M6	M7	M8	M9+	M10+				
		USA	Australia	Canada	Germany	UK	France	China	Spain	Netherlands	Russia			
Authorship (proportions)	USA	Black		Black									793	
	Australia		Black					Grey					193	
	Canada	Black		Black					Black				188	-6%
	Germany				Black								183	
	UK	Black		Black		Black							178	
	France					Black	Black						132	
	China	Grey						Black					115	
	Spain			Grey					Black				110	-6%
	Netherlands	Black								Black			81	
	Russia	Grey											32	
Readers		2860	224	588	487	736	335	109	245	75	6			

+ = NegBin replacing ZINB

Table 5. Education & Educational Research. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

n=2026		Location of readers										Papers	Author bias	
Model	M1	M2^	M3+	M4*	M5+	M6+	M7+	M8*	M9+	M10				
		USA	UK	Australia	Canada	Turkey	Taiwan	China	Netherlands	S. Africa	New Zealand			
Authorship (proportions)	USA	Black	Grey										693	2%
	UK	Grey	Black										279	
	Australia			Black									222	
	Canada	Grey			Black								118	
	Turkey		Grey			Black							97	-21%
	Taiwan						Black						95	
	China	Black											62	
	Netherlands		Black										55	
	S. Africa			Grey						Black			45	
	N. Zealand												43	
Readers		2264	948	448	278	64	15	28	121	132	42			

^ = simple ZINB replacing standard ZINB, * = ZIP replacing ZINB, + = NegBin replacing ZINB

Table 6. Nursing. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

n=1080		Location of readers										Papers	Author bias
Model	M1	M2+	M3	M4+	M5+	M6+	M7	M8+	M9+	M10+			
		USA	Australia	UK	Brazil	Sweden	Taiwan	Canada	Norway	China	S. Korea		
Authorship (proportions)	USA	█		█								415	9%
	Australia		█	█								125	
	UK			█	█							86	
	Brazil	█		█	█							70	
	Sweden											66	
	Taiwan	█										62	
	Canada	█					█					54	
	Norway							█				38	
	China											29	
	S. Korea											16	
Readers		492	42	254	125	12	2	145	7	1	3		

+=NegBin replacing ZINB

Table 7. Information Science & Library Science. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

n=664		Location of readers										Papers	Author bias
Model	M1	M2	M3+	M4	M5	M6+	M7+	M8+	M9	M10			
		USA	UK	Canada	Australia	Spain	Taiwan	China	S. Africa	Netherlands	Germany		
Authorship (proportions)	USA	█	█	█					█			264	-14%
	UK		█	█								59	6%
	Canada	█		█								47	8%
	Australia				█							30	-15%
	Spain					█						28	20%
	Taiwan						█					24	
	China							█				23	21%
	S. Africa								█			19	
	Netherl.	█										18	
	Germany										█	18	
Readers		945	269	169	50	196	3	14	35	72	169		

+=NegBin replacing ZINB

Table 8. Artificial Intelligence. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

n=624		Location of readers										Papers	Author bias
Model	M1	M2	M3	M4	M5	M6*	M7+	M8+	M9	M10+			
	China	USA	Germany	Spain	UK	France	Italy	S. Korea	Canada	Taiwan			
China												118	
USA												110	
Germany												52	
Spain												51	
UK												44	
France												42	
Italy												30	
S. Korea												24	
Canada												23	
Taiwan												19	
Readers	142	209	154	58	110	92	45	14	50	2			

*=ZIP replacing ZINB, +=NegBin replacing ZINB

Table 9. Mech Eng. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

n=352		Location of readers										Papers	Author bias
Model	M1+	M2+	M3	M4+	M5+	M6+	M7+	M8+	M9+	M10-			
	USA	China	UK	France	Italy	Canada	S. Korea	India	Iran	Taiwan			
USA												102	
China												48	
UK												31	
France												25	
Italy												22	
Canada												19	
S. Korea												12	
India												6	
Iran												6	
Taiwan												6	
Readers	129	19	108	42	15	9	7	18	26	0			

+=NegBin replacing ZINB, -=no model fitted

Table 10. Philosophy. Black (resp. grey) cells indicate that authorship from the country in the row is a significant predictor of increased (resp. decreased) readers from the country in the column. Each column is a ZINB model unless stated.

n=246		Location of readers										Papers	Author bias
		Model	M1+	M2+	M3+	M4+	M5+	M6+	M7+	M8+	M9+		
		USA	UK	Germany	Australia	Canada	Belgium	France	S. Africa	Italy	Spain		
Authorship (proportions)	USA	Black										123	
	UK		Black									39	
	Germany											15	
	Australia											13	
	Canada											12	
	Belgium											7	
	France											2	
	S. Africa											2	
	Italy											2	
	Spain											2	
	Readers	114	90	25	14	64	1	6	2	20	11		

+ = NegBin replacing ZINB

Discussion

The results show some clear but not universal trends. They should be interpreted cautiously, however, due to the limitations of the data. In particular, since a majority of Mendeley users do not record a country, most Mendeley users are excluded. In addition, most article readers do not use Mendeley, and the proportion of Mendeley readers is likely to vary by country, and so the results above, which are only about registered Mendeley users that record a country, are a significantly limited and nationally biased sample. In partial mitigation, however, it is not clear that biases in the sample would be towards excluding (or including) people that tend to read articles from their own country, especially because the sample is of English articles and the Mendeley interface is likely to be either in English or in the language of the user, so national languages should not be a factor.

In answer to the first research question, there is a tendency for articles to be more read in countries with a higher share of their authorship. Whilst the results are most clear cut for fields and countries with the most data, in general, there is a strong tendency for articles to have disproportionately many readers from countries that co-authored them. In the tables, this is evident from the black shading of the diagonal in many places (including all except two of the 40 diagonals in the largest data sets – tables 1-4) and the absence of grey squares. Whilst the shading pattern is not universal, this could be due to a lack of sufficient data (i.e., statistical power), and so it is plausible that the trend applies in most cases. Education is an exception, however, since this field has a relatively large amount of data but no real evidence of a diagonal trend. This is surprising because education research seems likely to have some national components, at least in part, due to the differing national education systems. Nevertheless, there are education research areas that are explicitly international (e.g., *Comparative Education* is indexed in WoS), and education students might be expected to research other education systems to shed light on their own by contrast or to get ideas of good practice to implement in a local context. There was clear evidence of international interest in national research since some of the papers had national titles but international readers. For example, the Slovakian-authored paper "Slovakian students' knowledge of and attitudes toward biotechnology" had readers from the USA, the UK and Mexico (but none from Slovakia), and the USA-authored paper "Madrasas and Pakistan's education agenda: Western media misrepresentation

and policy recommendations" had readers from Bangladesh, Malaysia, and the UK. Nevertheless, it does not seem feasible that these factors would be enough to account for the lack of a clear trend in the data. Another possibility is that readers get local educational information predominantly from sources outside WoS, such as national journals and professional magazines, but tend to get international information from WoS journals. For example, WoS seems to index relatively few specifically regional Education journals, with the exceptions being *English in Australia*, *Journal of Baltic Science Education* and perhaps *Phi Delta Kappan* (a professional magazine in the USA), although others have a regional name but seem to be international in focus (e.g., *Australian/Asia Pacific/British/Croatian Journal of Education*, and many similar titles). To further investigate, the tests were rerun for a larger sample, all Education & Education Research WoS articles from 2002-2013 with Mendeley readers (n=26,167). For this new sample, all diagonal entries were significant and positive (not shown). This suggests that there *is* a tendency for Mendeley users to read Education & Education Research articles from their own country, but that this tendency is only substantial enough to appear in a very large sample.

One possible explanation for the tendency of Mendeley users to read articles co-authored from their own country could be that the authors are often readers of their own articles in Mendeley but this does not seem to be the case. Few of the tests for a bias towards authorship categories for Mendeley readers (the last column in the tables) were significant and the biases found were not systematically in favour of the authorship categories. Hence, although some authors almost certainly register as Mendeley readers of their own articles, this does not seem to occur enough to affect the results. Alternatively, it could be offset by other phenomena, such as authors recommending their own articles to their students, but this would also give an increased national audience for articles.

Although a likely reason for the tendency of people to read articles from their own country is that they know or have heard of the authors and therefore pay more attention to the articles, other explanations are also possible. The pattern could be due to different levels of interest for specialisms within the selected fields, so that people tend to read articles from their own country because they tend to be within a specialism that their country is interested in. Alternatively, some of the authors may have instructed undergraduate or postgraduate students in their own classes to read their articles. It also seems possible that there is a level of unconscious or conscious national bias in the sense of deliberately ignoring articles from other nations or from other specific nations, or in the sense of believing that research from other countries is inferior. There is no evidence in the data for or against any of these hypotheses, however, although it has been known for a long time that there are systematic biases within science (e.g., Merton, 1968, 1988), including perhaps also for countries (Bonitz, Bruckner, & Scharnhorst, 1997).

In answer to the second research question, there are cases where articles authored by one country are less (or more) read by other specific countries. Whilst some of these cases could be dismissed as a statistical anomaly because no Bonferroni correction was applied to the data, there are some general trends that suggest tentative explanations. First, articles from the People's Republic of China tend to attract fewer readers from other countries in several fields and the same seems to be true to some extent for Japan, India, Italy and Spain. Possible explanations for these are that articles from these countries tend to be published in lower-quality journals within the sets analysed, the authors travel less and are therefore less well-known internationally, the research tends to address niche areas of less interest to the international community, the standard of English is lower, or that there is simple unconscious or conscious bias against research from these countries. Again, these are purely speculative and there is no specific evidence for or against these hypotheses.

A limitation of the results is that only ten fields were tested and each field was pruned of interdisciplinary articles so the same results may not be replicated for interdisciplinary research, although this seems unlikely even though there is a relationship between interdisciplinarity and co-authorship (Qin, Lancaster, & Allen, 1997). Similarly, although all articles were in English, it seems probable that international biases would be significantly greater for articles written in any other language. A statistical limitation is that the tests reported do not take into account the magnitude of

the effects, only whether they are statistically significant or not, and that the data is assumed to be independent but this is not completely true, for example for multiple articles in the data set from the same authors.

Conclusions

The results give clear evidence that Mendeley readers disproportionately select articles from their own countries in many different fields. Whilst this applies only to English WoS articles, which presumably form the core of each discipline, the same result would presumably apply to wider samples of academic journals because larger samples outside the core of science would be more likely to include national journals and journals in languages other than English. The results apply to enough fields and countries to suggest that the national preference found may be a general, although perhaps not universal, phenomenon. Presumably this is not Mendeley-specific and the underlying cause is that article readers, in general, tend to come disproportionately from the authors' countries. Since the fields analysed excluded those with any obvious national affiliations (e.g., agriculture, history, law, economics, politics, zoology), the finding is worrying from the perspective of science as a global and impartial quest for knowledge, because it suggests that people sometimes read articles for reasons other than that they are the most relevant or important. Nevertheless, the results could also be due to differing national levels of specialisation within fields, which would not be a problem. However, it seems likely that human factors play a role in the results, with people noticing papers more if they are written by people that they have heard of, revealing the human side of science.

The few cases of negative international preferences (e.g., against papers authored by the People's Republic of China in some cases) seem to be particularly worrying to science as a whole and unfair to the authors of the papers in question, especially as they have the extra hurdle of publishing in a foreign language. An important practical implication of the presence of any type of national bias is that researchers should guard against any such biases in their reading and ensure that they read the best and most relevant articles because this should logically help them to produce the best research.

A corollary of the greater national readership for articles is that the work of authors in nations with a small science output is likely to be undervalued by citation-based or readership-based metrics because of the small domestic audience for their results. This gives theoretical support for a previous call for scale-independent indicators for the scientific performance of nations (Katz, 2000) and helps to explain the previously-identified Matthew Effect for countries (Bonitz, Bruckner, & Scharnhorst, 1997).

A final important implication of the results is that more internationally collaborative research may appear to be better because it attracts a larger audience in each of the countries of the authors rather than because the research is better. This casts doubt on claims about the benefits of international research collaboration, unless it is accepted that publicity for research is a valuable outcome in itself. Hence, research funders and policy-makers should not justify incentivising international collaborative research on the basis that it is likely to be better than national research, although they may still wish to incentivise it on political grounds, for publicity, or for specific research needs, such as big science projects that individual countries could not easily fund.

Acknowledgment

This paper is supported by ACUMEN (Academic Careers Understood through Measurement and Norms) project, grant agreement number 266632, under the Seventh Framework Program of the European Union.

References

- Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*, 60(7), 1320-1326.
- 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.
- Bar-Ilan, J. (2012). JASIST@Mendeley. In ACM Web Science Conference 2012 Workshop. <http://altmetrics.org/altmetrics12/bar-ilan/>
- Bollen, J., & Sompel, H. V. D. (2008). Usage impact factor: The effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology*, 59(1), 136-149.
- Bonitz, M., Bruckner, E., & Scharnhorst, A. (1997). Characteristics and impact of the Matthew effect for countries. *Scientometrics*, 40(3), 407-422.
- Bote, G., Vicente, P., Olmeda-Gómez, C., & Moya-Anegón, F. (2013). Quantifying the benefits of international scientific collaboration. *Journal of the American Society for Information Science and Technology*, 64(2), 392-404.
- Cronin, B., Shaw, D., & Barre, K. L. (2004). Visible, less visible, and invisible work: Patterns of collaboration in 20th century chemistry. *Journal of the American Society for Information Science and Technology*, 55(2), 160-168.
- de Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., & Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53-78.
- Didegah, F., Thelwall, M., & Gazni, A. (2012). An international comparison of journal publishing and citing behaviours. *Journal of Informetrics*, 6(4), 516-531.
- Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, 7(4), 861-873.
- Duin, D., King, D., & van den Besselaar, P. (2012). Identifying audiences of e-infrastructures - tools for measuring impact. *PLoS ONE*, 7(12), e50943. doi:10.1371/journal.pone.0050943
- Francescheta, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics*, 4(4), 540-553.
- Gazni, A., Sugimoto, C. R., & Didegah, F. (2012). Mapping world scientific collaboration: authors, institutions, and countries. *Journal of the American Society for Information Science and Technology*, 63(2), 323-335.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge*. London, UK: Sage.
- Glänzel, W., & Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics*, 61(3), 395-404.
- Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69-115.
- Gunn, W. (2013). Social signals reflect academic impact: What it means when a scholar adds a paper to Mendeley. *Information Standards Quarterly*, 25(2), 33-39.
- Henning, V., & Reichelt, J. (2008). Mendeley - A Last.fm for research? In IEEE Fourth International Conference on eScience (eScience'08) (pp. 327-328). Los Alamitos: IEEE Press.
- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge: Cambridge University Press.
- Hoekman, J., Frenken, K., & Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, 39(5), 662-673.
- Jaffe, A. B., & Trajtenberg, M. (1999). International knowledge flows: Evidence from patent citations. *Economics of Innovation and New Technology*, 8(1-2), 105-136.
- Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 451-554.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration?. *Research policy*, 26(1), 1-18.

- Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31-43.
- Katz, J. S. (2000). Scale-independent indicators and research evaluation. *Science and Public Policy*, 27(1), 23-36.
- Kurtz, M. J., & Bollen, J. (2010). Usage bibliometrics. *Annual Review of Information Science and Technology*, 44, 1-64.
- Lancho-Barrantes, B. S., Bote, G., Vicente, P., Rodríguez, Z. C., & de Moya Anegón, F. (2012). Citation flows in the zones of influence of scientific collaborations. *Journal of the American Society for Information Science and Technology*, 63(3), 481-489.
- Lancho-Barrantes, B. S., Guerrero-Bote, V. P., & de Moya-Anegón, F. (2013). Citation increments between collaborating countries. *Scientometrics*, 94(3), 817-831.
- Lasthiotakis, H., Sigurdson, K., & Sá, C. M. (2013). Pursuing scientific excellence globally: Internationalising research as a policy target. *Journal of Higher Education Policy and Management*, 35(6), 612-625.
- Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91(2), 461-471.
- Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In É. Archambault, Y. Gingras & V. Larivière (Eds.), *17th international conference on science and technology indicators (STI 2012)* (pp. 541-551). Montréal: Science-Metrix and OST.
- Liang, L., Rousseau, R., & Zhong, Z. (2013). Non-English journals and papers in physics and chemistry: bias in citations? *Scientometrics*, 95(1), 333-350.
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology & Human Values*, 17(1), 101-126.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56-63.
- Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *ISIS*, 79(4), 606-623.
- Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the American Society for Information Science and Technology*, 56(10), 1088-1097.
- Mohammadi, E., & Thelwall, M. (in press). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the American Society for Information Science and Technology*.
- Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421-432.
- Persson, O. (2010). Are highly cited papers more international? *Scientometrics*, 83(2), 397-401.
- Price, D. J. d. S. (1963). *Little science, big science*. New York: Columbia University Press.
- Priem, J., & Hemminger, B. H. (2010). *Scientometrics 2.0: New metrics of scholarly impact on the social Web*. *First Monday*, 15(7). <http://firstmonday.org/ojs/index.php/fm/article/view/2874>
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2011). Altmetrics: A manifesto, Available: <http://altmetrics.org/manifesto/>.
- Qin, J., Lancaster, F. W., & Allen, B. (1997). Types and levels of collaboration in interdisciplinary research in the sciences. *Journal of the American Society for information Science*, 48(10), 893-916.
- Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2013). Download vs. citation vs. readership data: the case of an information systems journal. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (eds.), *Proceedings of the 14th International Society of Scientometrics and Informetrics Conference, Vienna, Austria* (pp. 626-634). Vienna: Facultas Verlags- und Buchhandels AG.
- Schlögl, C., & Gorraiz, J. (2011). Global usage versus global citation metrics: the case of pharmacology journals. *Journal of the American Society for Information Science and Technology*, 62(1), 161-170.

- Schwarz, A. W. (1999). Scientific centres in Europe: an analysis of research strength and patterns of specialisation based on bibliometric indicators. *Urban Studies*, 36(3), 453-477.
- Sud, P. & Thelwall, M. (submitted). Not all international collaboration is beneficial: The Mendeley readership and citation impact of biochemical research collaboration. <http://cybermetrics.wlv.ac.uk/NotAllInternationalCollaboration.pdf>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLOS ONE*, 8(5), e64841. doi:10.1371/journal.pone.0064841
- Torres, D., Cabezas, Á., & Jiménez, E. (2013). Altmetrics: New indicators for scientific communication in Web 2.0. *Comunicar*, 41(21), 53-60.
- Van Leeuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & Van Raan, A. F. (2011). First evidence of serious language-bias in the use of citation analysis for the evaluation of national science systems. *Research Evaluation*, 9(2), 155-156.
- Van Raan, A. F. J. (1998). The influence of international collaboration on the impact of research results. *Scientometrics*, 42(3), 423-428.
- Wallace, M. L., Larivière, V., & Gingras, Y. (2012). A small world of citations? The influence of collaboration networks on citation practices. *PloS ONE*, 7(3), e33339.
- Wuchty, S., Jones, B., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316, 1036-1038.
- Zahedi, Z., Costas, R., & Wouters, P. (2013a). How well developed are Altmetrics? Cross-disciplinary analysis of the presence of 'Alternative Metrics' in scientific publications. In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (eds.), Proceedings the 14th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2013), Vol. 1, Vienna: Facultas Verlags- und Buchhandels AG, (pp. 876-884).
- Zahedi, Z., Costas, R., & Wouters, P. (2013b). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? ALM Workshop, San Francisco, CA. <http://article-level-metrics.plos.org/files/2013/10/Zahedi.pptx>
- Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. *Tech Trends*, 55(1), 32-36.