

Contributorship and division of labor in knowledge production

Vincent Larivière

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

Observatoire des sciences et des technologies, (OST-CIRST), Université du Québec à
Montréal, Montréal, Qc, Canada

Nadine Desrochers

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

Benoît Macaluso

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

Observatoire des sciences et des technologies, (OST-CIRST), Université du Québec à
Montréal, Montréal, Qc, Canada

Philippe Mongeon

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

Adèle Paul-Hus

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

Cassidy R Sugimoto

School of Informatics and Computing, Indiana University Bloomington, Bloomington, IN,
USA

Abstract

Scientific authorship has been increasingly complemented with contributorship statements. While such statements are said to ensure more equitable credit and responsibility attribution, they also provide an opportunity to examine the roles and functions that authors play in the construction of knowledge and the relationship between these roles and authorship order. Drawing on a comprehensive and multidisciplinary dataset of 87,002 documents in which contributorship statements are found, this paper examines the forms that division of labor takes across disciplines, the relationships between various types of contributions, as well as the relationships between the contribution types and various indicators of authors' seniority. It shows that scientific work is more highly divided in medical disciplines than in mathematics, physics and disciplines of the social sciences, and that, with the exception of medicine, the writing of the paper is the task most often associated with authorship. The results suggest a clear distinction between contributions that could be labelled as 'technical'

and those that could be considered 'conceptual': While conceptual tasks are typically associated with authors with higher seniority, technical tasks are more often performed by younger scholars. Finally, results provide evidence of a u-shaped relationship between extent of contribution and author order: In all disciplines, first and last authors typically contribute to more tasks than middle authors. The paper concludes with a discussion of the implications of the results for the reward system of science.

Keywords

contributorship, authorship, collaboration, bibliometrics

Correspondence

Vincent Larivière Université de Montréal, École de bibliothéconomie et des sciences de l'information, C.P. 6128, Succ. Centre-Ville, H3C 3J7 Montréal, Qc., Canada

Email: vincent.lariviere@umontreal.ca

Introduction

Authorship allows for the attribution of credit for an idea or a discovery, assigns moral rights and responsibility (Biagioli, 1999; Birnholtz, 2006), and provides a basis for peer recognition (Merton, 1973) in a reputation economy in which symbolic capital is the main currency (Bourdieu, 1975, 2001).

The historical conception of an author has been challenged by the massive increase in co-authorship over the last few decades (Larivière et al., 2015). Cronin (2005: 50-51) has highlighted the fact that 'the practice of promiscuous co-authorship puts considerable stress on [the] tried and true model' upon which the relationship between symbolic capital and responsibility is built in the traditional reward system of science. Authorship in high energy physics (HEP) presents a stark example of the complexities of what Cronin (2005) terms 'hyperauthorship'. For example, a HEP publication with more than 5,000 authors was published in 2015, setting the record for the largest number of contributors to a single research article (Castelvecchi, 2015) and calling into question

the allocation of credit and responsibility for this work. The multiplication of the number of authors on publications has the double consequence of multiplying the credit attributed for the knowledge produced – each author can claim a publication count – and dividing the responsibility for its reliability.

The complexity of assessing contribution in the case of multi-authorship has been a long-standing area of investigation in the sociology of science. Researchers have debated the appropriate methods for counting authorship (e.g., straight vs. fractional counts) (Lindsey, 1980) and have examined deviant behavior in assigning authorship (e.g., Heffner, 1979). Rising co-authorship rates have been associated with an increase in unethical research practices (Bennett and Taylor, 2003; Wislar et al., 2011), particularly in the case of ‘guest authorship’ – where individuals who did not make a substantial contribution to a work are included as authors – and, inversely, ‘ghost authorship’ – where an individual who made a substantial contribution is excluded from the author list. These deviant practices are not particularly rare: analyses of the biomedical literature have shown evidence of honorific authorship in 18-19% of sampled articles and of ghost authorship in 8-11% of articles (Flanagin et al., 1998; Wislar et al., 2011).

Such ambiguity in assigning authorship makes it more difficult to hold scientists accountable, especially in the cases of scientific fraud (Biagoli, 1999; Garcia et al., 2010; Kevles, 1998; Wray, 2006) and leads to the sharing of responsibility for fraudulent work among all collaborators in cases of misconduct perpetuated by one colleague (Mongeon and Larivière, 2014). Given the consequences of fraud for biomedical research, the International Committee of Medical Journal Editors (ICMJE) responded to concerns in rising co-authorship rates with a statement on specific requirements for the attribution of authorship, first published in 1988. In the 2014 version of the *Recommendations for*

the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals, it is stated that authorship must be based on the following four criteria, all of which must be met:

- substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work, AND
- drafting the work or revising it critically for important intellectual content, AND
- final approval of the version to be published, AND
- agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. (ICMJE, 2014: 2)

In addition, the committee resolved that authors should not only be able to identify and vouch for their own contributions, but be able to identify the contributions of all co-authors and verify that the requirements for authorship have been met (ICMJE, 2014).

Despite the existence of these guidelines, authorship attribution in the medical fields does not always reflect substantial contributions (Sismondo, 2009). Furthermore, ICMJE recommendations are not able to deal with disciplinary differences in authorship practices (Biagioli, 2003; Birnholtz, 2006; Pontille, 2004). For example, authorship in HEP is seen in terms of 'credits for accumulated labor' (Biagioli, 2003: 270) and every contributor taking part in an experiment signs the paper, generally in alphabetical order. Examples from the social sciences and humanities (SSH) provided by Pontille (2004) suggest, on the other hand, that the writing of a paper is the key to authorship. That being said, quantitative SSH fields are more likely to provide collaboration opportunities than qualitative SSH fields (Moody, 2004) and might reward other type of contributions. In biology, first position is attributed to the individual who performed most of the

experiment and the last position is reserved for the head of the laboratory, who typically signs each scientific article produced by his or her lab; technicians, graduate students, and other researchers who contributed to the research to a lesser extent occupy middle positions (Pontille, 2004) – historically, they would often have been invisible (Shapin, 1989). Yet, the ICMJE guidelines do not grant authorship to people who contribute only such things as technical help, writing assistance, or general and monetary support.

In response to current transformations of scientific authorship practices, new forms of credit attribution have emerged. Contributorship, where a list of the specific contributions of each listed author is systematically recorded, was first proposed and adopted by biomedical journals in the late 1990s (Rennie et al., 1997; 2000; Smith, 1997; Topol, 1998). Still limited to a few journals – *Nature*, *PNAS*, the *British Medical Journal*, as well as the PLOS series of journals, among others – the contributorship model has been adopted primarily in biomedical and associated fields (Wager, 2007).

There have also been other fledgling attempts to document contributions to knowledge production: co-first authorship, as well as primary and secondary author hierarchies, are increasingly present in biomedical fields (Conte et al., 2013); authorship badges were recently discussed as a means to clarify the roles of co-authors (Singh Chawla, 2014); and digital humanities initiatives have sought various ways of modeling contributorship to take into account the new types of tasks associated with producing knowledge (e.g., the Taxonomy of Digital Research Activities in the Humanities, or TaDiRAH). Acknowledgments also provide information that can help to better understand scientific credit attribution practices, often revealing the role of individuals and organizations who contributed to a research project, but whose contributions did not warrant authorship status (Cronin et al., 2003; Desrochers et al., 2016).

Despite these advances in documenting contributorship, there have been no large-scale empirical studies examining the roles and functions that authors play in the construction of knowledge and the relationships between these roles and authorship order. This leaves scholars to speak in anecdotes or dated examples. Revealing the nature of the work that warrants authorship across domains and the relationship between this work and authorship order provides critical information for the allocation of reward and the assignment of responsibility in scientific communication. Therefore, the objective of this paper is to inform a more comprehensive understanding of the roles and functions of authorship in contemporary scientific activity in relation to the symbolic capital associated with authorship order. Taking disciplinary variation into account, this paper uses contributorship data found in PLOS articles to answer the following research questions:

- To what degree are authors sharing the responsibility of contributions?
- Which contribution types are likely to appear in isolation or in combination with other contributions?
- What is the relationship between academic age and type of contribution?
- How does type of contribution vary by author position?

Current knowledge on this topic rests on several assumptions; among these, that authors are observing guidelines about authorship and that authorship position serves as an adequate proxy for type and level of contribution to a scientific paper. However, these assumptions lack large-scale empirical testing. This paper provides an exploratory analysis of division of labor in knowledge production, and of how this division varies across disciplines.

Methods

Data sources

Two sources of data are used in this paper: 1) all articles published by the Public Library of Science (PLOS), available on the PLOS website in XML format; and 2) Thomson Reuters' Web of Science (WoS), including the Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI), and the Arts & Humanities Citation Index (AHCI). PLOS was chosen for the comprehensive data it provides, thanks to its policy on mandatory contributorship self-declarations. The first PLOS journal – *PLOS Biology* – was published in 2003. Between 2003 and 2014, 127,911 articles have been published in PLOS journals. *PLOS ONE* is the most prolific, with 106,460 documents published between 2006 and October 9, 2014 (the date of data collection). Combined, the seven other journals published a total of 21,451 documents.

Data collection

We used the DOI of each PLOS article to download the full text and match the article with its corresponding record in the WoS. We built a code to automatically download the XML format of each PLOS paper. Using the DOIs found in the Article-Level-Metrics table provided on the PLOS website (PLOS, n.d.), the URL of the XML format for each of the documents was built and queried using SQL Server Integration Services (SSIS) as well as a Visual C# code. This automatically uploads the XML format of each article to the user's computer.

Given that the metadata of each PLOS paper is available in the WoS in a much more usable format than in the PLOS XML, only two elements are needed from the PLOS articles' XML structure: the DOI and the list of authors' contributions. Articles' DOIs – obtained from their URLs – were used to match articles with the corresponding record

in the WoS. In order to isolate and retrieve the authors' contributions from the full text of the articles, another Visual C# script integrated to SSIS was written.

Table 1 provides the number of PLOS articles retrieved and the number of PLOS articles in the WoS, as well as the proportion of PLOS articles that were matched with the WoS. As shown, 97.6 % (n=94,879) of all PLOS articles were indexed in the WoS. However, given that not all PLOS articles were assigned a DOI in the WoS, there is a small fraction of PLOS articles that could not be matched. In total, more than 95.5% of PLOS articles published between 2008 and 2013 were matched to the WoS (n=92,845). Most of the articles were published in the journal *PLOS ONE* (more than 85%), and the large majority of these could be matched with the WoS (98.2%). The journal *PLOS Clinical Trials*, which was only published between 2006 and 2007, was excluded from the analysis, since it was not published during the period covered. The lower proportion of papers matched in *PLOS Biology* is due to a small number of papers indexed in the WoS.

Table 1. Number and percentage of papers published in PLOS journals indexed in the WoS, 2008-2013

Journal	PLoS	WoS	Direct link with DOI	Match (%)
PLOS Biology	1,693	1,380	988	58.4%
PLOS Computational Biology	2,569	2,429	2,090	81.4%
PLOS Genetics	3,456	3,251	2,865	82.9%
PLOS Medicine	1,275	1,214	975	76.5%
PLOS Neglected Tropical Diseases	2,346	2,197	2,098	89.4%
PLOS ONE	82,656	81,393	81,208	98.2%
PLOS Pathogens	3,214	3,015	2,621	81.5%
All PLOS Journals	97,209	94,879	92,845	95.5%

The dataset of PLOS articles, including authors' contributions as well as all WoS metadata, served as the sampling frame for the study. From this, only articles and review articles were included, given that these are peer-reviewed and more likely to represent

important contributions to knowledge (Moed, 2006). After removing articles lacking author contribution data (n=962), those for which a match could not be established between PLOS and WoS (n=369), and duplicates (n=4), the final dataset included 87,002 articles.

We isolated each contributing author and obtained contribution statements for all 87,002 articles. These articles contained 20,667 distinct contribution labels, associated with 40,356 initials (contributors), for more than 1.5 million records. After correcting typographical errors and grouping the contribution statements having the same meaning (for example, 'writing the paper' and 'writing the manuscript'), we obtained a list of the most common contributions, as well as the number of articles and of author-article combinations (or authorships) that feature each contribution. Table 2 presents the proportion of papers in which at least one author performed a given contribution, as well as the overall proportion of authorships (i.e., sum of all author-paper combinations) for each contribution. It shows that almost all papers were signed by authors who analyzed the data, conceived and designed the experiments, performed the experiments, and wrote the paper, and that about 75% of papers had at least one author who contributed reagents, materials or tools of analysis.

Table 2. Number and percentage of articles and of authorships (author-paper combinations), by type of contribution

Contribution	Articles		Authorships	
	N	%	N	%
Analyzed the data	85,900	98.7%	320,080	50.6%
Conceived and designed the experiments	85,406	98.2%	288,765	45.6%
Contributed reagents/materials/analysis tools	64,444	74.1%	220,331	34.8%
Performed the experiments	82,811	95.2%	311,679	49.3%
Wrote the paper	86,517	99.4%	287,796	45.5%
<i>Other (20 243)</i>	<i>15,900</i>	<i>18.3%</i>	<i>79,978</i>	<i>12.6%</i>
N distinct papers	87,002	100.0%	632,799	100.0%

Attribution of a discipline to PLOS articles

In order to analyze disciplinary differences in the division of scientific labor, we needed to assign a discipline to each article. Drawing on the method developed by Waltman and van Eck (2012), we categorized each article into one discipline of the National Science Foundation classification scheme (NSF, 2006) on the basis of its citation and referencing patterns. More specifically, we compiled, for each article, the distribution, by discipline, of references made and citations received, and assigned to the paper the discipline for which the sum of these two proportions was the highest. As expected, PLOS articles are dispersed in varied disciplines, but mainly biomedical research (44.3%), clinical medicine (43.5%), biology (7.3%), and psychology (1.8%). All other disciplines account for less than 1% of papers.

Estimation of academic age

Academic age of authors was estimated using year of first publication, as recorded in the WoS. In order to obtain said age, authors found in the WoS were disambiguated automatically by the Center for Science and Technology Studies (CWTS, Leiden University) using the algorithm developed by Caron and van Eck (2014). The use of first

publication year has been shown to be a reasonable predictor of researchers' biological age and year of PhD (Costas et al., 2015). In this paper, results are presented for researchers who have an academic age between 0 (i.e., first year in which a contributor was listed on a paper) and 30 (i.e., 30 years since their first paper).

Results

The (un)evenness of contributions

The first concept we measured was the evenness of contributions across disciplines – that is, the degree to which most or all of the authors made all five types of contributions. More specifically, we identified the fields in which most authors contributed to a majority of the tasks and those where research activity was more fragmented. Figure 1 presents the distribution of authors as a function of their number of contributions [1-5].

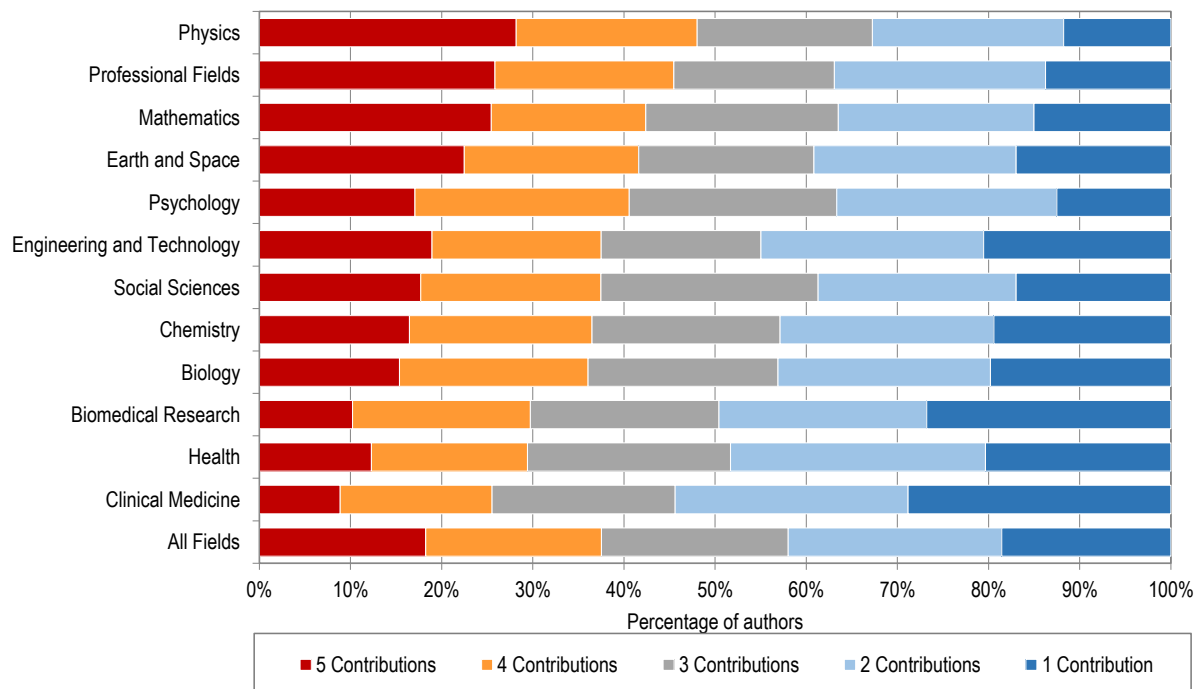


Figure 1. Distribution of authors as a function of their number of contributions, by discipline

Figure 1 illustrates that biomedical research and clinical medicine are disciplines in which the contribution is more fragmented, with 10% or less of all authors having contributed to all five tasks. Conversely, these are also the two disciplines with the highest proportion (27%-29%) of authors who contributed to one task only. At the other end of the spectrum, mathematics, physics, and professional fields are disciplines where the largest proportion of authors contributed to all five task ($\approx 25\%$), and where the lowest proportion of authors contributed to one task. Between these two extremes, disciplines such as psychology, health, engineering and technology, chemistry, and biology have a mode (the number of contributions with the highest frequency) of two, while social sciences have a mode of three.

Tasks leading to authorship

In order to provide a better understanding of the various tasks that are associated with authorship, we compiled the mean proportion of authors per paper who have contributed to each specific task, by discipline (Figure 2). As one might expect from Figure 1, biomedical research and clinical medicine exhibit a much greater division of labor, with most tasks being performed by roughly 50% to 60% of authors. For these two disciplines, as well as for chemistry, the contribution that is most often associated with authorship is *Analyzed the data*, while the least common task is *Contributed reagents/materials/analysis tools*. For all of the nine other disciplines, *Writing the paper* is the most common contribution, with percentages around 70% for biology, engineering and technology, earth and space, and health (social sciences), and percentages around 80% for psychology, social sciences, mathematics, professional fields and physics. This suggests that writing the paper is a quasi-essential criterion for the attribution of authorship in these fields. Also, in this group of nine disciplines,

Performing the experiments remains associated with a lower proportion of authors – below 60% – except in social sciences, where it is higher. This suggests that in those disciplines, manual experimental work is much less likely to be associated with authorship than the *Conception and design of the experiments* and the *Analysis of the data*, which are performed by a higher proportion of authors. This general dichotomy between biomedical research, clinical medicine and chemistry, on the one hand, and the nine other disciplines, on the other, could indicate that the first three are more likely to associate technical and experimental work with authorship, while the nine others are more likely to associate authorship with contributions considered more conceptual and intellectual.

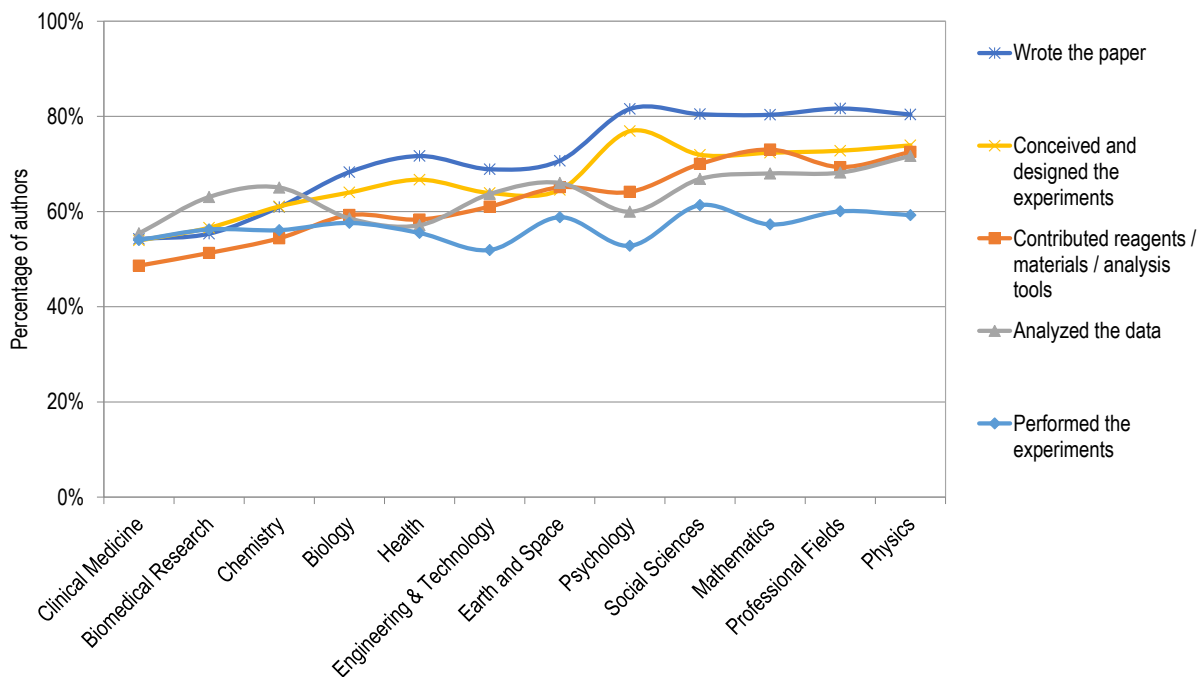


Figure 2. Percentage of authors who contributed to a specific task, by discipline

Coupling and isolation of types of contributions

The ICMJE recommendation suggests that authors should be associated with all key tasks; however, as implied by the previous tables, that is rarely the case. Typically, certain tasks are isolated and others are coupled. Figure 3 provides a heat map

demonstrating the association between the number of contributions an author makes and type of contribution. As shown, when an author contributes to only one task, this is most likely to be performing an experiment and, to a lesser extent, contributing materials. An author contributing in only one way is unlikely to contribute by writing the paper or designing the experiment. On the other hand, those who contribute to three or four tasks are much less likely to have *Performed the experiments* or *Contributed reagents/materials/analysis tools*. This suggests a clear distinction between contributions that could be labelled as ‘technical’ and those that could be labelled as ‘conceptual’ (the next section provides more details on the relationship between such contributions and the seniority of authors).

Contribution	Nb. of Contribution				
	1	2	3	4	5
Analyzed the data	Blue	White	Red	Red	White
Conceived and designed the experiments	Blue	White	Red	Red	White
Contributed reagents/materials/analysis tools	Red	Blue	Blue	Blue	White
Performed the experiments	Red	White	Blue	White	White
Wrote the paper	Blue	White	Red	Red	White

Figure 3. Association between the number of contributions an author makes and the type of contribution (red denotes a strong association, blue indicates a lack of association, and white indicates neutrality)

Appendix A details these relationships by discipline. Biology, biomedical research, chemistry, and clinical medicine share patterns of behavior in regards to the relative isolation of *Performed the experiments* and *Contributed reagents/materials/analysis tools*. These patterns are present, but less pronounced, in earth and space, engineering, and technology, health, and the social sciences. A shift is seen in mathematics, physics,

and professional fields, where *Wrote the paper* becomes comparatively more isolated than *Performed the experiments* (i.e., those only contributing in one way are more likely to be contributing to writing than experimentation).

Certain tasks are also likely to appear in concert. Figure 4 denotes the association between, for example, those who analyze the data and those who contribute materials. As shown, *Wrote the paper* is highly associated with *Analyzed the data* and *Conceived and designed the experiments*, suggesting that these activities are typically performed in tandem. *Performed the experiments*, when not in isolation, is likely to be done by those who also *Analyzed the data*.

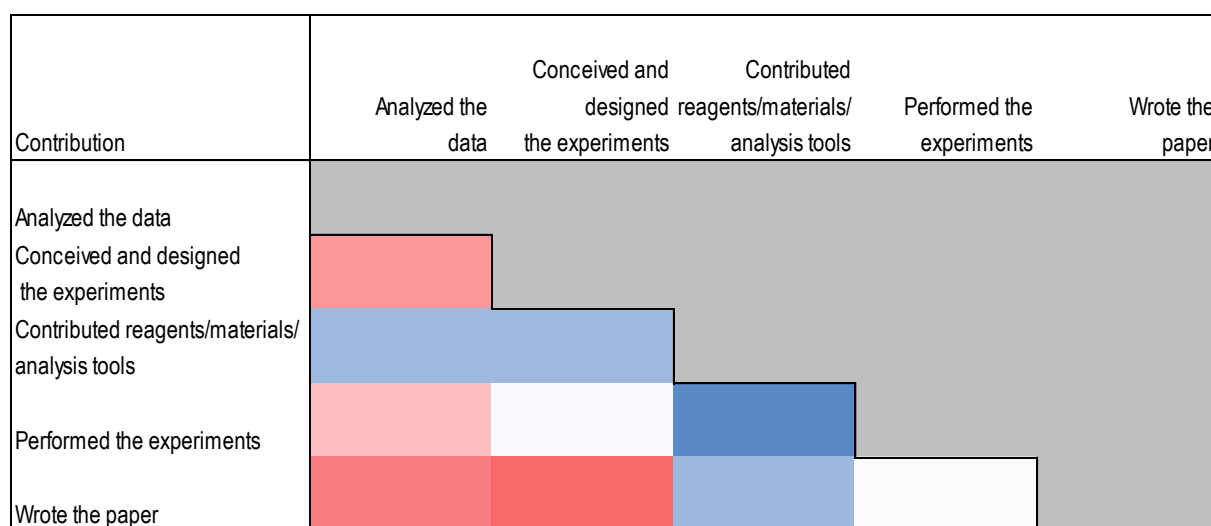


Figure 4. Association between contributions at the author level (red denotes a strong association, blue indicates a lack of association, and white indicates neutrality)

Academic age and type of contribution

One could expect that authors' academic age affects how they contribute to knowledge production. Table 3 demonstrates the varying ages associated with certain contributions: specifically, younger scholars are more likely to be contribute to the

experimentation (mean age = 6.8 years since first publication), while mid-career scholars (i.e., those with around 10 years of academic age) show a stronger association with *Analyzing the data*. *Writing the paper*, *contributing reagents*, and *conceiving and designing the experiments* are tasks associated with scholars with more than 12 years since their first publication.

Table 3. Mean academic age of authors (time since first publication) associated with contribution type

Contribution	Mean age
Conceived and designed the experiments	13.2
Contributed reagents/materials/analysis tools	12.7
Wrote the paper	12.6
Analyzed the data	10.5
Performed the experiments	6.8

Figure 5 depicts the evolution of these contribution types. While *Performed the experiments* is the contribution most likely associated with authors who are, on average, 7 years and younger, it is the contribution less likely to be done by authors who are, on average, 14 years and older. This suggests a shift from technical work to more conceptual work as researchers age and rise in the hierarchy of science. Interestingly, we do not observe such a decrease for *Analyzed the data*: while this task is also more likely to be performed by young authors, it remains a common contribution as researchers get older.

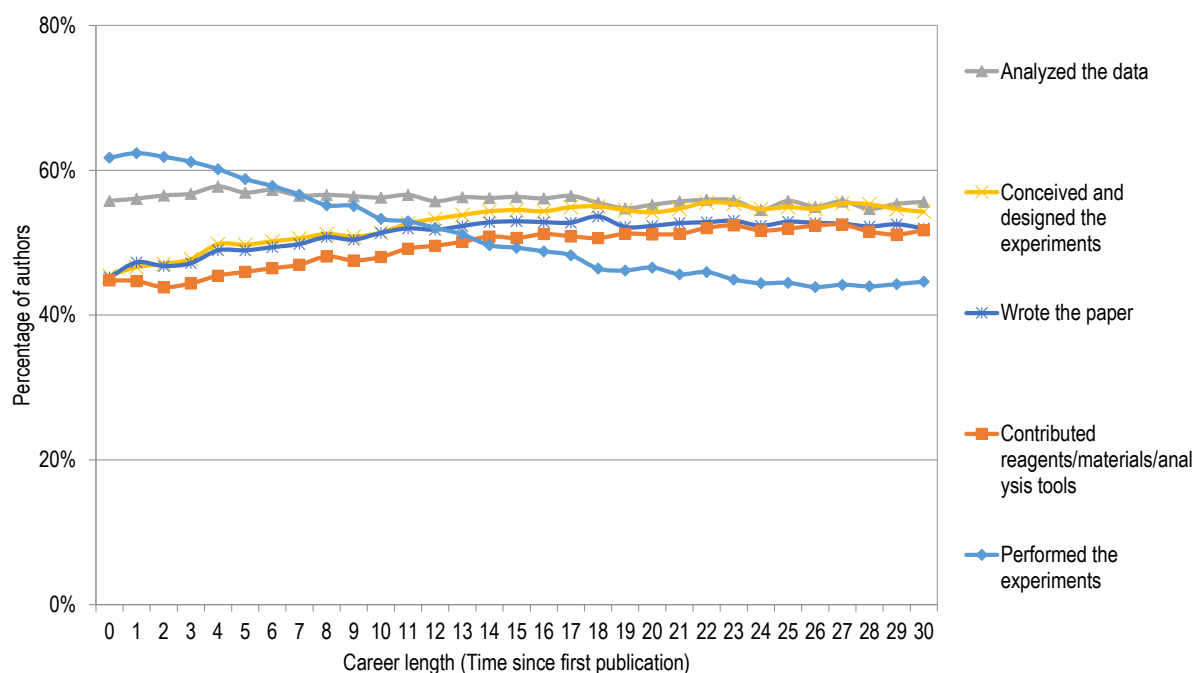


Figure 5. Percentage of authors who have contributed to a specific task, by academic age

Contributorship and author position

In many fields, authorship order is used as a proxy to indicate dominant positions, the first and last position being the most commonly used. One relationship between authors' order and contribution can be seen in the proportion of tasks to which authors of the various author positions have contributed. Figure 6 shows that, in every discipline, first authors are contributing to a higher proportion of tasks than middle and last authors. More specifically, first authors were associated with more than 82% of the tasks – with a few disciplines having percentages above 90% – while last authors are in the 60%-68% range. For middle authors, these percentages vary dramatically, from 42% of contributions in biomedical research to 59% in professional fields. On the whole, this figure provides evidence of a U-shaped structure in the relationship between author's order and contribution – in which the most important authors of a paper are the first and last authors, with a lower contribution from authors in middle positions.

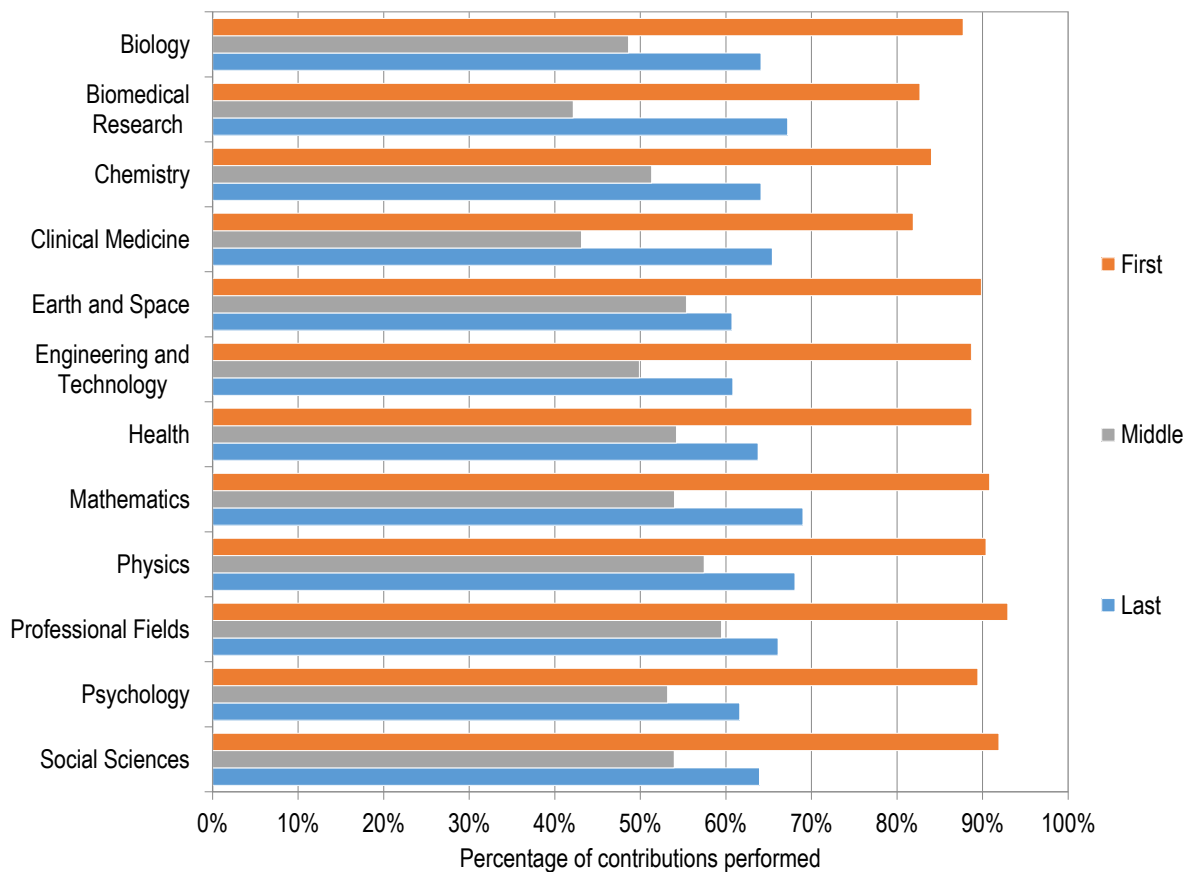


Figure 6. Percentage of contributions performed as a function of authors' order, by discipline

Discussion

Evenness and distribution of contributions in the various disciplines

Although collaboration has risen in all disciplines (Larivière et al., 2016), the data reveals a resilience of the scientific habitus, which serves as a 'social conservation instinct' (Bourdieu, 1988: 150) that supports certain behaviors and disciplinary traditions. This has a direct influence on the relationship between contributorship and authorship. For example, the relatively lower level of collaboration in mathematics compared with other disciplines of the sciences (Grossman, 2002) – a likely consequence of the theoretical nature of the work – explains the evenness of contributions that lead to authorship in this field: A very high proportion of authors

have contributed to all of the tasks involved. On the other hand, the evenness of trends found in physics is not based on low levels of collaboration or division of labor but, rather, on the egalitarian nature of the field, found in the standard practice, at least in high energy physics, of alphabetically listing authors and the established generosity in authorship attribution (Biagioli, 1999; Birnholtz, 2006). In other words, in the case of physics, contributorship statements are more likely to be a reflection of physicists' adherence to the Mertonian ideal of communalism (Merton, 1973) than to practices of equal contribution.

These forms of habitus are key to understanding how the distribution of contributions are directly linked to the *illusio* – the set of rules that shapes both the game and the stakes for all players (Bourdieu, 1988: 56). Given that adherence to this *illusio* is a *sine qua non* of participation in the field, and since this *illusio* is the very foundation upon which symbolic capital is obtained, it goes without saying that scientists will protect the traditions of the dissemination of knowledge in their field even as the field changes. Therefore, while contributorship statements call for the standardization of authorship attribution, the fact that these statements remain self-declared – like author lists and acknowledgments – calls for contextualization.

Relationship between contribution type and authorship across disciplines

The data presented here suggest that work is not distributed equally, challenging the recommendations that authors be 'accountable for *all aspects of the work* in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved' (ICMJE, 2014, emphasis added). The 'whole-to-part' and 'part-to-whole' relationships described (one might say required) in the guidelines do not reflect the reality of research. Collaborators are not involved in all

aspects of the work and those conducting the experiments appear to be the most isolated. This seems to contravene the first ICMJE criterion, according to which an author must have made 'substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work' (ICMJE, 2014). Furthermore, as the numbers of collaborators on each paper skyrockets, the feasibility of a system where one is both accountable for and able to 'identify which co-authors are responsible for other specific parts of the work' (ICMJE, 2014) is diminished. In this context, contributorship statements might increase transparency in research by holding individuals accountable for stated contributions. It remains to be seen, however, whether such statements do make scholars more accountable in cases of scientific fraud and, more generally, given the precise assignation of tasks that they provide, whether they have a dissuasive effect on researchers' propensity to commit misconduct.

Relationship between academic age and type of contribution

The age of researchers and the nature of their contributions are also clearly related. Older researchers (12.6 years from first paper) are most involved in three types of contributions: design, writing, and contribution of materials. *Conceived and designed the experiments* (mean number of years: 13.2) is akin to an intellectual seal, a claim of paternity on the discovery, innovation, or conception; without this, none of the other steps would have been possible. It therefore consecrates the researcher as leader amongst peers and is likely to be associated with those in power positions, such as the principal investigator (PI). Along the same lines, *Wrote the paper* (12.6 years) has traditionally been considered the most direct claim to authorship, both as the recognition of the credit it brings and as an acceptance of the responsibility it conveys (Biagioli, 1999). While this can be shared, its value in the reward system of science makes it very difficult to renounce, however advanced one's academic career might be.

The third type of contribution, however, can be associated with a different type of recognition. *Contributed reagents/materials/analysis tools* (mean years: 12.7) denotes access to means. The capital here is, therefore, mainly financial, and typically obtained through reputation and recognition – hence Merton’s ‘principle of cumulative advantage’ (Merton, 1973: 457-458). The age gap between this task and others is exacerbated by the increasing gap between available positions and the number of PhDs granted (Schillebeeckx et al., 2013) and the changing demographics of the scientific workforce (Arbesman and Wray, 2013), which has led to an increase in the mean age at which scholars obtain their first grant (National Institute on Aging, 2015).

The difference between the mean age of authors who have *Analyzed the data* (10.5 years) and those who have *Performed the experiments* (6.8 years) suggests that the latter task is quickly abandoned by scholars after their initial, legitimating years, roughly corresponding to promotion and tenure in many academic systems. Performing the analysis means getting one’s scientific hands dirty, but it also means having an intimate knowledge of the data. It is therefore a type of contribution that is within reach of younger scientists, but one that scholars continue to tend to for a longer period of time than performing the experiments. It is also closely linked to the responsibility aspect of authorship, which is pivotal in the scientific field, since analysis is often where flaws can be discovered and where the heart of the work’s contribution to the world of knowledge is situated.

Positional authority

This paper demonstrates that first and last authors are contributing to a higher proportion of tasks than middle authors, with first authors contributing to more tasks – typically all of the tasks – than last authors. These dominant authors are more likely to

be associated with conceptual tasks, while technical tasks are conducted by middle authors. Thus, that which was documented at the micro-level by Pontille (2004) for lab-based disciplines can be observed across the entire spectrum of subfields, even in the social sciences. Despite the existence of these two poles of dominance (i.e., first and last authorship) – which are associated with different levels of seniority and contributions – in each disciplines the first author is associated with a larger proportion of the contributions and generally can be considered the principal or lead author of the research paper.

Limitations and further research

Just as acknowledgments have been both described as potential indicators and criticized for their self-declared, subjective nature and lack of standardization (Cronin, 1992; Cronin and Weaver-Wozniak, 1993; Desrochers et al., 2015; McCain, 1991), contributorship statements are not yet a common source of data for understanding the relationships between authorship and knowledge production. They are not indexed in large-scale bibliographic databases and are only required by a handful of journals; partly for these reasons, this research focuses primarily on the natural and medical sciences, which have been the first disciplines to require contributorship statements. Broader analyses, including that of the full-text of papers, would be useful in understanding the extent to which these contribution statements reflect the construction of the papers and whether the papers support the idea of compartmentalized contributions or make it more difficult to pinpoint who did what, in spite of these statements. In this context, insights on scholars' perception of contributorship statements and on the level of 'seriousness' with which they are made would help to contextualize the results presented here. Along these lines, disciplinary practices are crucial to this topic. In furthering this research, it would be interesting to look at multi- or interdisciplinary

teams, where the validation of each contributor's methods and tools by all other authors is simply not feasible, the strength of the teams being built precisely on the difference of their approaches and training. The same could be said of national and international collaborations.

Conclusion

Rennie and colleagues (1997) introduced the contributorship model as a radical alternative to the traditional authorship model in which the concept of contributor would replace the concept of author, positing that '[a]bandoning the concept of author in favor of contributor frees us from the historical and emotional connotations of authorship, and leads us to a concept that is far more in line with the actuality of modern scientific cooperative work' (Rennie et al, 1997: 582). The objective of that proposition was pragmatic: to ensure a more equitable credit attribution and, more importantly, guarantee a precise and reliable responsibility attribution. However, almost a decade later, less than 10% of a broad sample of biomedical journals included a contributors list (Wager, 2007). Though many institutions and research councils require that faculty provide narratives of their contribution to multi-authored work when for review in promotion, tenure and funding decisions, to date, contributorship has not been formally incorporated into research assessment exercises. This is largely due to a lack of systematic indexing and limited adoption by journals and publishers. In the current scholarly communication system, authors, contributors, and acknowledgees coexist, sharing diverse, complementary, and sometimes redundant functions that remain to be better understood.

As the first large-scale analysis of contributorship, this work has shed light on the degree to which contributorship can inform our understanding of the practice of science and

the allocation of rewards and resources within it. Contributorship is particularly important given the rising fragmentation of tasks in team science. In a manner similar to authorship – albeit more precise – it can be used to credit authors for specific tasks, but also, in the case of scientific misconduct, to assign responsibility. As such, contributorship statements are not merely a reaction to hyperauthorship. Rather, they provide greater transparency in the scientific process and its division of labor, and aim at holding scholars accountable for their work. Time will tell whether such functions are seen as sufficiently important by the research community to make such statements mandatory.

Acknowledgments

The authors wish to thank Rodrigo Costas and Ludo Waltman (CWTS-Leiden University) for providing the data on the year of first publication, Elise Smith for stimulating discussions, as well as the anonymous referees for their precise and constructive feedback.

Cited References

- Arbesman S and Wray K (2013) Demographics and the fate of the young scientist. *Social Studies of Science* 43(2): 282-286
- Bennett DM and Taylor DM (2003) Unethical practices in authorship of scientific papers. *Emergency Medicine* 15(3): 263-270.
- Biagioli M (1998) The instability of authorship: Credit and responsibility in contemporary biomedicine. *The FASEB Journal* 12(1): 3-16.
- Biagioli M (1999) Aporias of scientific authorship: Credit and responsibility in contemporary biomedicine. In: Biagioli M (ed), *The Science Studies Reader*. London: Routledge, 12-31.
- Biagioli M (2003) Rights or rewards? Changing frameworks of scientific authorship. In: Galison P and Biagioli M (eds), *Scientific Authorship : Credit and Intellectual Property in Science*. New York, NY: Routledge, 255-279.
- Birnholtz JP (2006) What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology* 57(13): 1758-1770.
- Bourdieu P (1975) The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information* 14(6): 19-47.
- Bourdieu P (1988) *Homo Academicus*. Stanford, CA: Stanford University Press.
- Bourdieu P (2001) *Science de la Science et Réflexivité: Cours du Collège de France, 2000-2001*. Paris: Raisons d'agir.
- Caron E and van Eck NJ (2014) Large scale author name disambiguation using rule-based scoring and clustering. In: *19th International Conference on Science and Technology Indicators. 'Context Counts: Pathways to Master Big Data and Little Data'*. Leiden, the Netherlands, 79-86.
- Castelvecchi D (2015) Physics paper sets record with more than 5,000 authors. *Nature News* (15 May). Available at: <http://www.nature.com/news/physics-paper-sets-record-with-more-than-5-000-authors-1.17567> (accessed 24 April, 2016).
- Conte ML, Maat SL and Omary MB (2013) Increased co-first authorships in biomedical and clinical publications: A call for recognition. *The FASEB Journal* 27(10): 3902-3904.
- Costas R, Nane T and Larivière V (2015) Is the year of first publication a good proxy of scholars' academic age? In: Salah AA, Tonta Y, Akdağ Salah AA, et al. (eds), *Proceedings of the 15th International Conference on Scientometrics and Informetrics*. Istanbul, Turkey, 988-998.
- Cronin B (1992) Acknowledged but ignored: Credit where credit's due. *Bulletin of the American Society for Information Science* 18(3): 25.

- Cronin B (2005) *The Hand of Science: Academic Writing and its Rewards*. Lanham, MD: Scarecrow Press.
- Cronin, B and Weaver-Wozniak S (1993) Online access to acknowledgements. In: Williams, M.E. (ed), *Proceedings of the 14th National Online Meeting 1993*, Medford, NJ: Learned Information, 93-98
- Cronin B, Shaw D and La Barre K (2003) A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology* 54(9): 855–871.
- Desrochers N, Paul-Hus A and Pecoskie J (2015) Founding concepts and foundational work: Establishing the framework for the use of acknowledgments as indicators. In: Salah AA, Tonta Y, Akdağ Salah AA, et al. (eds), *Proceedings of the 15th International Conference on Scientometrics and Informetrics*, Istanbul, Turkey, pp. 890–894.
- Desrochers N, Paul-Hus A and Larivière V (2016) The angle sum theory: Exploring the literature on acknowledgments in scholarly communication. In: Sugimoto CR (ed), *Theories of Informetrics and Scholarly Communication*, De Gruyter Mouton, pp. 225–247.
- Flanagin A, Carey LA, Fontanarosa PB, et al. (1998) Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *JAMA* 280(3): 222–224.
- Garcia CC, Nabuco Martrucelli CR, Freire Rossilho M de M, et al. (2010) Authorship for scientific papers: The new challenges. *Revista Brasileira De Cirurgia Cardiovascular* 25(4): 559–567.
- Grossman JW (2002) Patterns of collaboration in mathematical research. *SIAM News* 35(9): 8-9.
- Heffner AG (1979) Authorship recognition of subordinates in collaborative research. *Social Studies of Science* 9(3): 377–384.
- International Committee of Medical Journal Editors (2014) Recommendations. Available at: <http://www.icmje.org/icmje-recommendations.pdf> (accessed 1 August 2014).
- Kevles DJ (1998) *The Baltimore Case: A Trial of Politics, Science, and Character*. New York: Norton.
- Larivière V, Gingras Y, Sugimoto CR, et al. (2015) Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology* 66(7): 1323–1332.
- Lindsey D (1980) Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science* 10(2): 145–162.

- McCain KW (1991) Communication, competition, and secrecy: The production and dissemination of research-related information in genetics. *Science, Technology & Human Values* 16(4): 491–516.
- Merton RK (1973) *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago and London: Chicago University Press.
- Moed HF (2005) *Citation Analysis in Research Evaluation*. Dordrecht ; Great Britain: Springer.
- Mongeon P and Larivière V (2015) Costly collaborations: The impact of scientific fraud on co-authors' careers. *Journal of the Association for Information Science and Technology* 67(3): 535-542.
- Moody J (2004) The Structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review* 69(2): 213–238.
- National Institute on Aging (2015) R01 teams and grantee age trends in grant funding. *Inside NIA: A Blog for Researchers*. Available at: <https://www.nia.nih.gov/research/blog/2015/04/r01-teams-and-grantee-age-trends-grant-funding> (accessed 24 April, 2016).
- National Science Foundation (2006). *Science and Engineering Indicators*. Chapter 5: Academic Research and Development. Data and Terminology. Available at: <http://www.nsf.gov/statistics/seind06/c5/c5s3.htm#sb1>.
- PLOS (n.d.) PLOS ALM data. Available at: <http://article-level-metrics.PLOS.org/PLOS-alm-data/> (accessed 24 April, 2016).
- Pontille D (2004) *La Signature Scientifique : Une Sociologie Pragmatique de l'Attribution*. CNRS Sociologie, Paris: CNRS.
- Rennie D, Yank V and Emanuel L (1997) When authorship fails: A proposal to make contributors accountable. *JAMA* 278(7): 579–585.
- Rennie D, Flanagan A and Yank V (2000) The contributions of authors. *JAMA* 284(1): 89–91.
- Schillebeeckx M, Maricque B, and Lewis C (2013) The missing piece to changing the university culture. *Nature biotechnology* 31(10): 938-941.
- Singh Chawla D (2014) Could digital badges clarify the roles of co-authors? *Science* (3 Nov). Available at: <http://www.sciencemag.org/news/2014/11/could-digital-badges-clarify-roles-co-authors> (accessed 24 April, 2016).
- Sismondo S (2009) Ghosts in the machine: Publication planning in the medical sciences. *Social Studies of Science* 39(2): 171–198.
- Smith R (1997) Authorship is dying: Long live contributorship. *BMJ* 315(7110): 696.
- Topol EJ (1998) Drafter and draftees. *The Lancet* 352(9131): 897–898.

Wager E (2007) Do medical journals provide clear and consistent guidelines on authorship? *Medscape General Medicine* 9(3): 16.

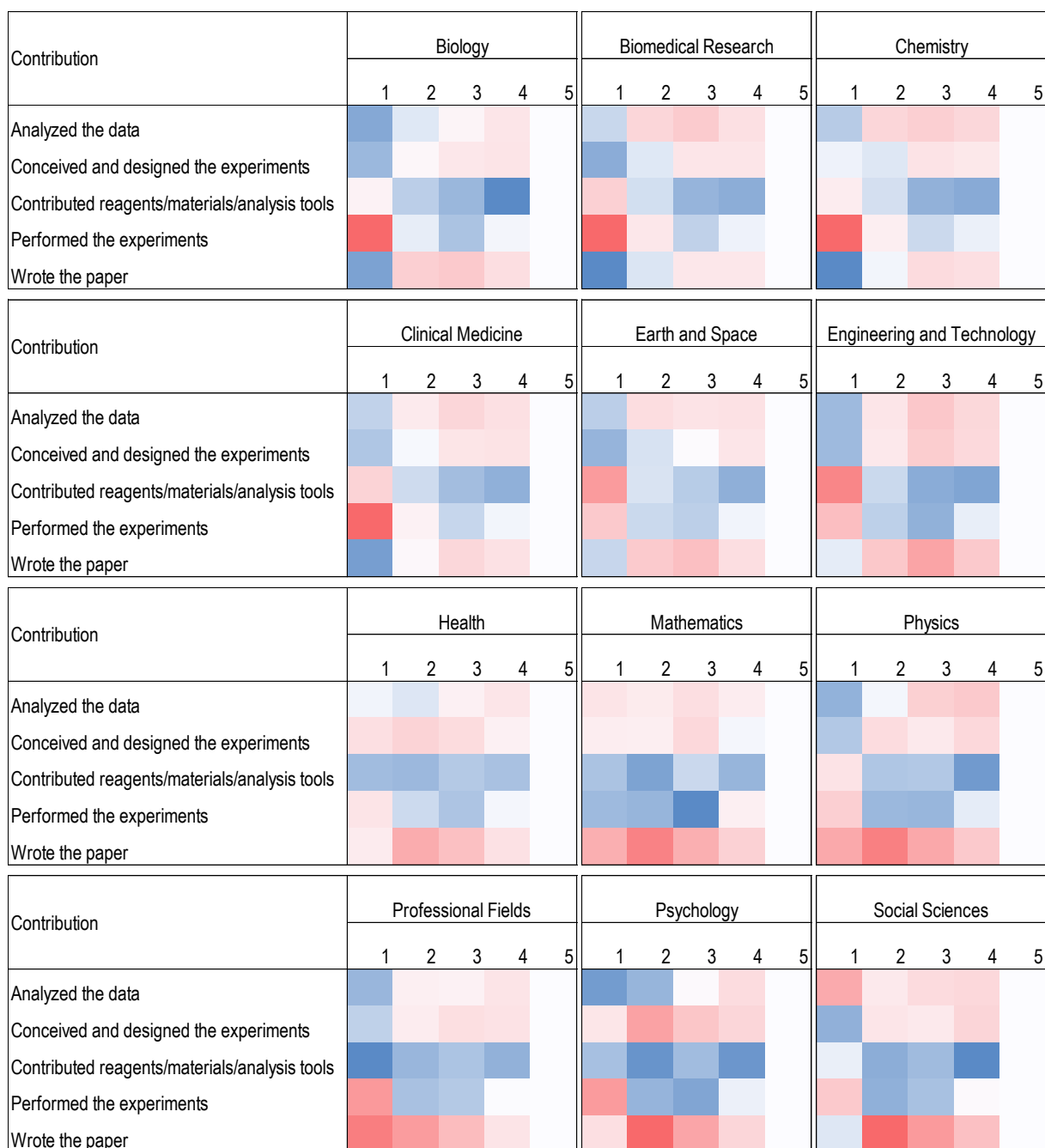
Waltman L and van Eck NJ (2012) A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology* 63(12): 2378–2392.

Wardil L and Hauert C (2015) Cooperation and coauthorship in scientific publishing. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 91(1-1): 012825.

Wislar JS, Flanagin A, Fontanarosa PB, et al. (2011) Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *BMJ* 343: d6128.

Wray KB (2006) Scientific authorship in the age of collaborative research. *Studies in History and Philosophy of Science Part A* 37(3): 505–514.

Appendix A. Association between the number of contributions an author makes and the type of contribution (red denotes a strong association, blue indicates a lack of association, and white indicates neutrality)



Author biographies

Vincent Larivière holds the Canada Research Chair on the Transformations of Scholarly Communication at the Université de Montréal. He is also the scientific director of the *Érudit* journal platform, and associate scientific director of the Observatoire des sciences et des technologies (CIRST-UQAM).

Nadine Desrochers holds degrees from Western University and the University of Ottawa. She is Assistant Professor at the École de bibliothéconomie et des sciences de l'information, Université de Montréal. Her research focuses on paratextuality and information behavior.

Benoît Macaluso is staff researcher at the Observatoire des sciences et des technologies (UQAM). He holds a master degree in information science from the Université de Montréal.

Philippe Mongeon is PhD candidate at the École de bibliothéconomie et des sciences de l'information, Université de Montréal. His research focuses on authorship and inventorship practices.

Adèle Paul-Hus is PhD candidate in information science at the École de bibliothéconomie et des sciences de l'information, Université de Montréal. Her doctoral research focuses on the functions of acknowledgements in the scholarly communication process.

Cassidy R. Sugimoto is Associate Professor in the School of Informatics and Computing at Indiana University Bloomington and President of the International Society for Scientometrics and Informetrics.