

Measuring disagreement in science

Dakota Murray¹, Wout Lamers², Kevin Boyack³, Vincent Larivière⁴, Cassidy R. Sugimoto¹,
Nees Jan van Eck², Ludo Waltman²

¹ dakmurra@iu.edu; sugimoto@indiana.edu

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

² w.s.lamers@cwts.leidenuniv.nl; ecknjpv@cwts.leidenuniv.nl; waltmanlr@cwts.leidenuniv.nl
Centre for Science and Technology Studies, Leiden University, Leiden, Netherlands

³ kboyack@mapofscience.com
SciTech Strategies, Inc., Albuquerque, USA

⁴ vincent.lariviere@umontreal.ca
École de bibliothéconomie et des sciences de l'information, Université de Montréal, Canada

Abstract

Dispute in science is central to the production of new knowledge. Such disputes leave traces in scholarly documents, generally through the form that is taken by citations. Based on the full text of scholarly papers from the Elsevier ScienceDirect database published between 1980 and 2016, this paper develops a methodology for investigating disagreement in science. Several signal phrases of disagreement are tested, and two are used (“contradict” and “conflict”, with the filter phrases “studies” or “results”) to assess the prevalence of disagreement across position within a paper and across disciplines. Results show that disagreement is relatively more common in the introduction and discussion sections of papers, as well as in fields of biomedical sciences, health sciences, social sciences, and humanities.

Introduction

Scientific disputes are central to the creation of new knowledge. More than 350 years ago, Robert Boyle and Thomas Hobbes debated the meaning of experimental results produced using Boyle’s newly-created air pump; from this controversy emerged the basis of modern scientific research (Shapin & Schaffer, 1985). Scholars have long been interested in studying controversy as it relates to the production of knowledge at the individual level (Latour, 1988), and the macro-development of science (Kuhn, 1962) and, often, make explicit norms that remain otherwise implicit (Gingras, 2014). More recently, doubts over scientific findings, such as for climate change research, has led scholars to measuring the degree of consensus of specific research areas (e.g., Oreskes, 2004; Shwed & Bearman, 2010). Information scientists have also studied disagreements among scientific literature, leveraging bibliometric tools to understand the development of scientific fields (Evans, 2007), characterize their differences (Fanelli & Glänzel, 2013), predict future scientific impact (Radicchi, 2012), measure uncertainty surrounding scientific claims (Chen et al., 2018), and to classify the function of citations (Catalini, Lacetera, & Oettl, 2015; Moravcsik & Murugesan, 1975).

This paper uses the full-text of scholarly publications to explore the degree of *controversy*, *disagreement*, and *dissonance* (henceforth referred to only as *disagreement*) in scientific literature. We examine sentences containing citations and identify a set of cue phrases that broadly signal disagreement between citing and cited paper, or within the cited literature. We assess the reliability of these cue phrases and use the top performing phrases to identify instances of disagreement in citing papers. This analysis provides a preliminary analysis of the degree of disagreement within fields. We also hope to establish a methodological basis for

future analyses of the disciplinary, temporal, and spatial aspects of scientific controversies through the lens of textual analysis.

Operationalizing disagreement

We used a broad operationalization of disagreement between a citing and cited paper, or within two cited papers. Under our definition, we consider such disagreement to include direct contradiction between conclusions, as well as disagreement based on incompatible model assumptions (even if findings are not in conflict). Examples of the types of disagreement we consider are shown in table 1.

Table 1. Examples of our notion of disagreement

<i>Citation sentence</i>	<i>Type of disagreement</i>
<i>A: "Coffee causes cancer"</i> <i>B: "Coffee does not cause cancer"</i>	Direct disagreement in conclusions
<i>C: "Based on a model which assumes that coffee increases the probability of cancer by 50%, the predicted life expectancy for the Dutch population equals 80 years."</i> <i>D: "Based on a model which assumes that coffee does not cause cancer, the predicted life expectancy for the Dutch population equals 85 years."</i>	Disagreement as a result of incompatible model assumptions, not necessarily because of conclusions
<i>E: "There remains controversy in the scientific literature over whether or not coffee is associated with an increased risk of cancer (A, B, C, D)"</i>	Disagreement in the broader literature

The main challenge is to obtain accurate signals of disagreement. For this purpose, we focus on sentences that include a citation, and that include a word or sequence of words signaling disagreement. We refer to this sequence of words as a *disagreement signal phrase*. In addition, other words appearing near this phrase may reinforce the likelihood that a disagreement signal phrase represents true disagreement—we call such words *disagreement filter phrases*

Data

We used data from the Elsevier ScienceDirect database hosted at the *Centre for Science and Technology Studies* at Leiden University. This data contains the full-text of nearly five million English-language research articles, short communications, and review articles published between 1980 and 2016. Sentences containing in-text citations are extracted from the full-text of these articles following the procedure outlined by Boyack et al. (2018).

Reliability

We considered four disagreement signal phrases: *contradict*, *contrast*, *conflict*, and *differ*. Queries include morphological variants of disagreement signal phrases, such that we include terms such as "conflict", "conflicted", and "conflicting". For each term, we used them as a standalone term (with no additional filters applied), and with one of four disagreement filter phrases: "ideas", "methods", "studies", and "results". Disagreement filter phrases must appear within a four-word window of the signal.

For each combination of disagreement signal and filter phrase, we randomly sampled 100 citation sentences from the full-text database that contain the combination of terms. Disagreement filter phrases must occur within a four-word window of the corresponding signal phrase. For each set of 100 sentences, two independent coders assigned a value of *valid*, or *invalid*, where valid means that the sentence represents a true example of our notion of

disagreement. In some cases, the proportion of *valid* instances was so low that coders did not code all 100 instances. Consider for example the four sentences listed below: the first is invalid, because the signal term, “conflict”, refers to an object of study, and not a scientific dispute; the second sentence is also invalid because the term “conflicting” refers to results within a single study, not between studies; the third and fourth sentence are both examples of sentences that would be marked as valid.

1. **Invalid:** “To facilitate conflict management and analysis in Mcr (...), the Graph Model for Conflict Resolution (GMCR) (...) was used.”
2. **Invalid:** “The 4-year extension study provided ambiguous [...] and conflicting post hoc [...] results.”
3. **Valid:** “These observations are rather in contradiction with Smith et al.’s [...].”
4. **Valid:** “Although there is substantial evidence supporting this idea, there are also recent conflicting reports (...)”

The validity—the proportion of sentences coded by both reviewers, and identified as *valid* by both reviewers—was calculated for each query (Figure 1). We find that the best performing disagreement signal phrases are “contradict” and “conflict”, and that these perform best when they occur alongside the disagreement filter phrases “studies” or “results”.

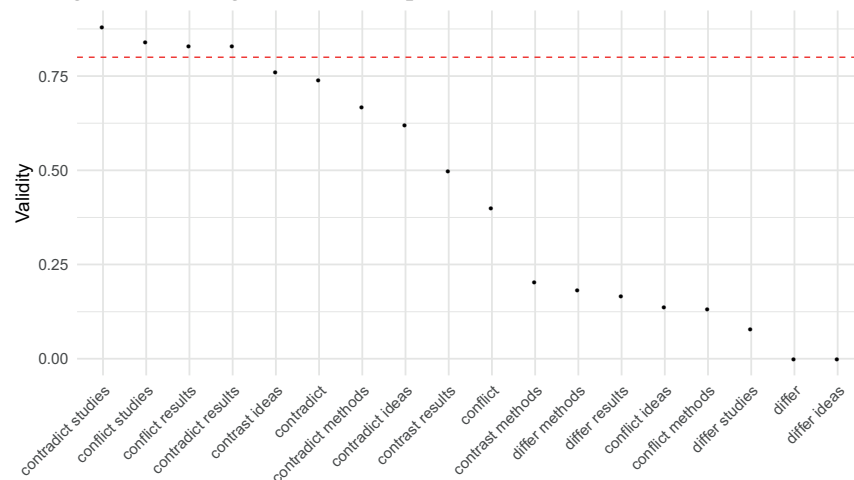


Figure 1: Validity of eighteen combinations of cue and signal words. We plot a threshold horizontal line of 0.80, showing the cut-off for choosing the top performing terms.

Analysis of “conflict” and “contradict” sentences

Due to their high validity, we focus our analysis on citation sentences containing the disagreement signal phrases “conflict” or “contradict”, which occur alongside filter phrases “studies” or “results”. There are, respectively, 62,667 and 63,035 “conflict” and “contradict” sentences in the text of our set of publications, each representing 0.04% of all citing sentences. During the period 1998–2016, the percentage of citing sentences having “conflict” or “contradict” has remained fairly stable over time. Below, we first report an analysis of the location of “conflict” and “contradict” sentences within the full text of publications. We then present a disciplinary comparison in which we examine the distribution of “conflict” and “contradict” sentences across scientific fields.

Location in full text

Figure 2 shows the distribution of “conflict” and “contradict” sentences within the full text of publications. The horizontal axis indicates text progression, expressed relatively to the total length of the full text of a publication. The vertical axis indicates the number of “conflict” or “contradict” sentences in a specific part of the full text of a publication relative to the total number of “conflict” or “contradict” sentences in the entire full text of a publication. The figure also shows the distribution of all citing sentences.

Consistent with earlier work (Bertin et al., 2016; Boyack, Van Eck, Colavizza, & Waltman, 2018), citing sentences are overrepresented in the early and to a lesser extent, the end parts of publications. For “conflict” and “contradict” sentences, this pattern is more pronounced. “contradict” sentences are especially overrepresented at the end of a publication, though less so in the early sections. In biomedical publications, the discussion of related work is often presented in the conclusion, thus leading to a large number of citing sentences at the end of publications (Boyack et al., 2018). As we will see below, “conflict” and “contradict” sentences occur most often in biomedical literature; potentially explaining why these sentences are overrepresented towards the ends of publications.

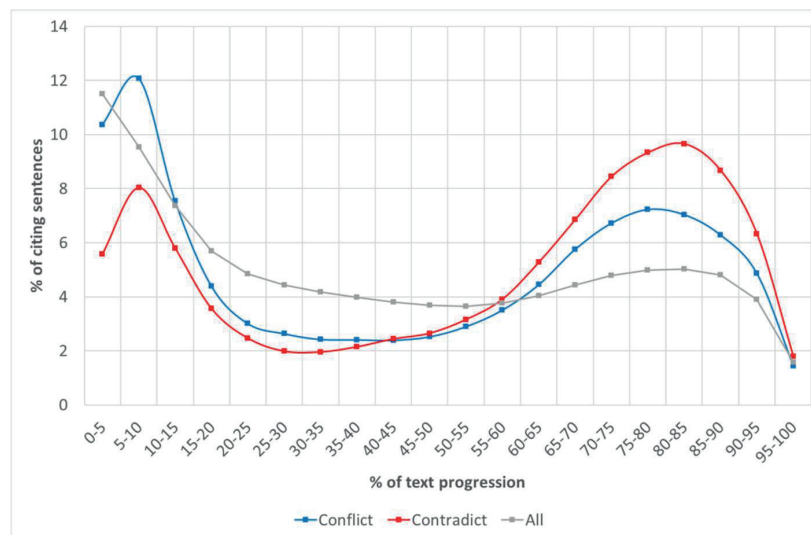


Figure 2. Distribution of “conflict” and “contradict” sentences within the full text of publications.

Disciplinary comparison

Our disciplinary comparison relies on all 2000-2017 publications indexed in the Web of Science, which were clustered into 868 fields through citation links, following the methodology introduced by Waltman and Van Eck (2012). For each field, we queried our Elsevier corpus and counted the total number of citing sentences, as well as the number of “conflict” and “contradict” sentences. Figure 3 presents visualizations of the 868 fields (nodes), produced using the VOSviewer software (Waltman & Van Eck, 2012). The size of a field indicates the total number of citing sentences in the field. The distance between two fields reflects the relatedness of the fields in terms of citation links: the smaller the distance between two fields, the larger the number of citation links between publications in the two fields. Most importantly, the color of a field indicates the relative number of “conflict” or “contradict” sentences in the field, expressed as the binary logarithm of the ratio of the actual and the expected number of “conflict” or “contradict” sentences. A field is colored blue if the number of “conflict” or “contradict” sentences is lower than the expected value, grey if it equals the expected value, and red if it is above the expected level.

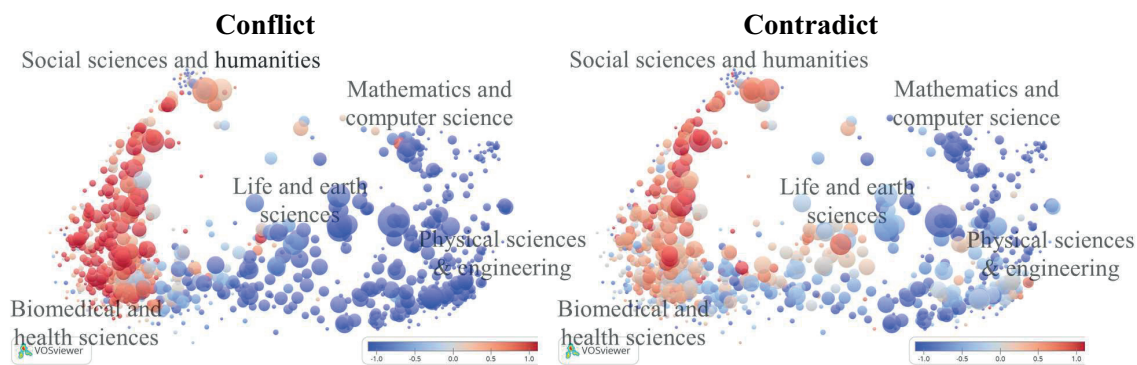


Figure 3. Distribution of “conflict” (left) and “contradict” (right) sentences over fields.

“Conflict” sentences were strongly concentrated in the biomedical and health sciences and in certain fields in the social sciences and humanities (roughly the top left and bottom left of each visualization); this was in strong contrast to the physical sciences, computer science, and mathematics (roughly top right and bottom right). In many of these fields, the number of “conflict” sentences was twice or more below expectation, whereas in the biomedical and health sciences, social sciences, and humanities, the number of conflict sentences was more than twice what was expected.

A similar, though less pronounced, trend was apparent for “contradict” sentences, with smaller disciplinary differences than for “conflict” sentences. In most biomedical and health fields, the number of “contradict” sentences was above expectation, but there were also some fields that fell below the expected number. Conversely, there were some fields in the life, earth, and physical sciences in which the number of “contradict” sentences was above expectation.

Table 2 lists the top 5 fields with the largest relative number of “conflict” and “contradict” sentences. We manually labelled each field by examining the titles of the journals with the largest number of publications in the field. The field labelled *International relations* had the largest relative number of “conflict” sentences. However, this was a methodological artefact. *International relations* studies political conflicts, and the term “conflict” referred mainly to conflicts as an object of study rather than conflicts in the scientific literature. Leaving out this field, all fields listed in table 2 were in the biomedical sciences and in psychology.

Table 2. Top 5 fields with the largest relative number of “conflict” and “contradict” sentences.

Conflict			Contradict		
Label	Absolute	Relative	Label	Absolute	Relative
International relations	414	3.11	Bioelectromagnetics	91	2.09
Cancer	67	3.04	Laboratory animals	45	1.71
Sleep medicine	229	2.26	Child psychology	89	1.50
Cardiothoracic surgery	143	2.25	Psychological methods	74	1.49
Cardiology	681	2.22	Cancer	23	1.48

Conclusion

This exploratory study assessed the degree to which *disagreement* between scientific literature exists across scientific fields. We defined a novel indicator of disagreement, and assessed the validity of a set of cue phrases that indicate disagreement between a citing and cited paper, or within the literature cited in a paper. We identified all citation sentences in our dataset that contained one of the two cue phrases (“contradict” and “conflict”, with the filter phrases

“studies” or “results”). Using these data, we investigated how the incidence of disagreement signal phrases differed based on their position in papers, noting key differences between sentences containing “contradict” or “conflict” signal phrases. We also investigated the incidence of disagreement signal phrases across 868 scientific fields represented by the Web of Science. We observed that the number of citing sentences containing disagreement signal phrases occurred above expected levels in the biomedical sciences, health sciences, social sciences, and humanities, and less than expected in the fields of mathematics, computer sciences, and physical sciences; we also noted that disciplinary differences were more extreme for “conflict” than for “contradict” sentences. Finally, we found that the fields with the largest proportion of conflict and contradict sentences were in the biomedical sciences and psychology.

This study marks the first step in an investigation into how disagreement and controversy function in scientific discourse. In future work, we will refine our notion of disagreement and expand our analysis to include additional disagreement signal and filter phrases. Building on this method, we hope to further investigate the extent to which disagreement and controversy relate to scientific impact; the evolution of the incidence of disagreement over time; how disagreement varies according to the country and institution of affiliation; and the incidence of disagreement as a function of the demographic characteristics of authors.

References

- Bertin, M., Atanassova, I., Gingras, Y., & Larivière, V. (2016). The invariant distribution of references in scientific articles. *Journal of the Association for Information Science and Technology*, 67(1), 164-177.
- Boyack, K. W., van Eck, N. J., Colavizza, G., & Waltman, L. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. *Journal of Informetrics*, 12(1), 59–73.
- Catalini, C., Lacetera, N., & Oettl, A. (2015). The incidence and role of negative citations in science. *Proceedings of the National Academy of Sciences of the United States of America*, 112(45), 13823–13826.
- Chen, C., Song, M., & Heo, G. E. (2018). A scalable and adaptive method for finding semantically equivalent cue words of uncertainty. *Journal of Informetrics*, 12(1), 158–180.
- Evans, J. H. (2007). Consensus and knowledge production in an academic field. *Poetics*, 35(1), 1–21.
- Fanelli, D., & Glänzel, W. (2013). Bibliometric Evidence for a Hierarchy of the Sciences. *PLOS ONE*, 8(6), e66938.
- Gingras, Y. (Editor) (2014) *Controversies. Accords et désaccords en sciences humaines et sociales*, Paris: CNRS. 278 p
- Latour, B. (1988). *Science in Action: How to Follow Scientists and Engineers Through Society* (Reprint edition). Cambridge, Mass: Harvard University Press.
- Moravcsik, M. J., & Murugesan, P. (1975). Some Results on the Function and Quality of Citations. *Social Studies of Science*, 5(1), 86–92.
- Oreskes, N. (2004). The Scientific Consensus on Climate Change. *Science*, 306(5702), 1686–1686.
- Radicchi, F. (2012). In science “there is no bad publicity”: Papers criticized in comments have high scientific impact. *Scientific Reports*, 2.
- Shapin, S., & Schaffer, S. (2011). *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (With a New introduction by the authors edition). Princeton, N.J: Princeton University Press.
- Shwed, U., & Bearman, P. S. (2010). The Temporal Structure of Scientific Consensus Formation. *American Sociological Review*, 75(6), 817–840.
- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.
- Waltman, L., & Eck, N. J. van. (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.