**Conceptualization**

We aim to perform an exploratory study of the lack of consensus in the scientific literature. A more precise elaboration of our concept of interest still needs to be performed, and ‘lack of consensus’ probably is not the most appropriate terminology. Possible alternatives for instance include ‘disagreement’ and ‘dissonance’. A review of the literature may yield additional suggestions on terminology. In what follows, I will use the term ‘disagreement’.

Importantly, we are not interested in disagreement within a single publication, but in disagreement between publications.

Disagreement between publications can be viewed in different ways. A narrow understanding of the concept of disagreement between publications is that two publications A and B are in disagreement only if the publications contain contradictory scientific propositions. For example:

* Publication A: Smoking causes cancer.
* Publication B: Smoking does not cause cancer.

A broader understanding of the concept of disagreement is that two publications A and B are in disagreement if publication A offers a perspective that is (at least partly) incompatible with the perspective offered by publication B (even though there does not need to be a contradiction between the scientific propositions contained in the two publications). For example:

* Publication A: Based on a model which assumes that smoking increases the probability of cancer by 50%, the predicted life expectancy for the Dutch population equals 80 years.
* Publication B: Based on a model which assumes that smoking does not cause cancer, the predicted life expectancy for the Dutch population equals 85 years.

There is no contradiction between the scientific propositions in publications A and B. However, the two publications use models that are based on incompatible assumptions. From a broader understanding of the concept of disagreement, we therefore consider the two publications to be in disagreement.

Another example:

* Publication A: Our data suggests that smoking increases the probability of cancer by 50%.
* Publication B: Our data does not provide evidence to support the hypothesis that smoking causes cancer.

If publications A and B use different data sets, there is no contradiction between the scientific propositions in the two publications. Nevertheless, from a broader understanding of the concept of disagreement, we do consider the two publications to be in disagreement.

The focus of our study will be on the broader understanding of the concept of disagreement as outlined above.

**Operationalization**

Our operationalization focuses on disagreement between publications that are related to each other by a direct citation link or a by co-citation link (at the sentence level). Disagreement between publications that do not have such a relation is not considered. Hence, if two communities of researchers that never cite each other reach contradictory conclusions on a certain issue, this will not be detected in our operationalization of disagreement.

To detect disagreement between publications, we examine the full text of publications and identify sentences that meet the following two conditions:

1. The sentence includes one or more references.
2. The sentence provides a semantic signal of disagreement.

Let’s consider a publication A, and let’s assume this publication includes a sentence that meets the above two conditions. The first condition is needed to ensure that the sentence does not just indicate internal disagreement within publication A. Suppose that the sentence includes references to publications B, C, and D. The semantic signal in the second condition may then indicate either disagreement between publication A on the one hand and publications B, C, and D (or at least some of these publications) on the other hand. Alternatively, the semantic signal may indicate disagreement among publications B, C, and D. In our operationalization of disagreement, we do not distinguish between these two possibilities. (This is an important difference with studies of negative citations, which focus exclusively on the first possibility.)

The main challenge is to obtain accurate signals of disagreement. For this purpose, we search for sentences that include a word or sequence of words signaling disagreement, referred to as a disagreement signal phrase. In addition, we may require that sentences also include another word or sequence of words, referred to as a disagreement filter phrase, that reinforces the likelihood that a disagreement signal phrase indeed refers to disagreement in the literature. The disagreement filter phrase must be located close to the disagreement signal phrase. There may be at most four words between the disagreement filter phrase and the disagreement signal phrase.

Examples of disagreement signal phrases include:

* contradiction
* contrast
* conflict
* controversy
* contrary

Some disagreement signal phrases have restrictions to avoid excessive numbers of false positives. For instance, ‘contrary’ has the restriction that it should not be part of the phrase ‘on the contrary’.

Examples of disagreement filter phrases include:

* result
* finding
* conclusion
* study
* literature

Disagreement filter phrases usually refer to objects about which there may be disagreement, such as results, findings, conclusions, or studies.

We aim to disregard negations. We therefore disregard sentences in which a negation word is located close to a disagreement signal phrase. A negation word is considered to be located close to a disagreement signal phrase if there are at most two words between the negation word and the disagreement signal phrase. We consider the following negation words:

* cannot
* neither
* no
* nor
* not

Closely related disagreement signal phrases may be grouped into a disagreement signal category. For instance, ‘disagree’, ‘not agree’, and ‘no agreement’ constitute a disagreement signal category. Likewise, closely related disagreement filter phrases may be grouped into a disagreement filter category. For instance, ‘approach’, ‘method’, ‘model’, and ‘technique’ constitute a disagreement filter category.

In order to obtain accurate signals of disagreement in a reasonably systematic way, we analyze each pair of a disagreement signal category and a disagreement filter category. The results of this analysis can be summarized in a matrix as shown below. The cells colored yellow indicate that a combination of a disagreement signal category and a disagreement filter category is considered sufficiently accurate to be used for providing a semantic signal of disagreement. For some disagreement signal categories, such as ‘controversy’, the cell in the ‘overall’ column is colored yellow. This indicates that the disagreement signal category on its own is sufficiently accurate to serve as a signal of disagreement. No additional disagreement filter category is needed.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Signal phrase** | **Result type** | **1**  **studies** | **2**  **results** | **3**  **methods** | **4**  **ideas** | **1234**  **total** | **5**  **standalone** |
| challenge\* | Sentences | 17108 | 16484 | 14525 | 7582 | 51536 | 427720 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| contradict\* | Sentences | 20374 | 54800 | 2900 | 5710 | 71510 | 122378 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| contrast\* | Sentences | 121198 | 123714 | 38455 | 8014 | 272949 | 1313707 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| contrary | Sentences | 17861 | 28196 | 4815 | 3802 | 51245 | 177821 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| conflict\* | Sentences | 23259 | 52616 | 5724 | 3728 | 73494 | 222215 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| disagree\* | Sentences | 5951 | 12953 | 1749 | 1196 | 19538 | 55535 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| differ\* | Sentences | 104179 | 114539 | 86982 | 9805 | 297922 | 2063891 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| controvers\* | Sentences | 12957 | 16458 | 3253 | 1973 | 32204 | 169801 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| disprov\* | Sentences | 303 | 393 | 107 | 587 | 1216 | 3200 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| refut\* | Sentences | 1565 | 2499 | 361 | 1692 | 5220 | 11401 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| debat\* | Sentences | 9001 | 5062 | 2848 | 1892 | 18223 | 162015 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| no consensus | Sentences | 1493 | 452 | 879 | 40 | 2777 | 17617 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |
| questionable | Sentences | 1149 | 2194 | 1245 | 895 | 5135 | 25985 |
| Accuracy |  |  |  |  |  |  |
| Kappa |  |  |  |  |  |  |