

A DEFINITELY NOT CHERRY-PICKED RHETORICAL ANALYSIS OF PROGRAMMING LANGUAGES REVIEWS

Hannah G. Ringler¹

Stefan Muller²

Abstract?

No, it's actually pretty concrete.

Introduction

Many academic papers receive multiple reviews, numbered according to some domain-specific scheme and labeled, e.g. “Review #1”, “Review #2”, etc. In such situations, it is well-documented that Reviewer #2 is the worst³⁴⁵. Some submission systems such as HotCRP⁶, however, identify reviews with letters (e.g. “Review A”, “Review B”, etc.) Since this identification scheme is relatively recent⁷, academics have had less time to draw broad generalizations about lettered reviews. A direct mapping between letters and numbers would lead us to believe that Reviewer B would display the same spiteful lack of empathy for which Reviewer 2 is known, but there is no inherent reason to believe that such a direct correlation exists. In this paper, we aim to ascertain, with mild amounts of accuracy, appropriate stereotypes of lettered reviews. Our “scientific” results show that, contrary to our assumption, it is Reviewer C who exists solely to crush the dreams of young researchers and not Reviewer B.

Methods

For our data, we gathered four sets of reviews from POPL and PLDI, received in 2017 and 2018 by member(s) of the Carnegie Mellon University community. To lend this study some semblance of scientific merit, we limited consideration to papers which received exactly four reviews, lettered A through D. Thus, we have a total of sixteen reviews. For each of these, we looked at only the pros, cons, comments, and questions, choosing to ignore the “objective” summaries that each reviewer provides. This allowed us to focus more completely on each reviewer’s actual feedback.

To analyze the reviews, we first used a rhetorical analysis tool entitled Docuscope⁸, which creates statistical counts of different “types” of language that are present in each of the texts. From these counts, we calculated averages of each type of language, and then returned to

¹ English Department, Carnegie Mellon University

² Computer Science Department, Carnegie Mellon University

³<https://www.facebook.com/academicssay/photos/a.1459750144246778.1073741828.1452615238293602/2066556870232766/?type=3&theater>

⁴ <https://www.facebook.com/groups/reviewer2/about/>

⁵ Anonymous, “Ode to Reviewer Two,” in *Proceedings of the tenth annual intercalary robot dance party in celebration of workshop on symposium about 2^{6th} birthdays*; in particular that of Harry Q Bovik (SIGBOVIK ’16), 2016, p.

0x8bcdcd109b835f1d5d612f3a9d451a2de98839a3c0a9548dd937e43ebdaa6436e.

⁶ Eddie Kohler, HotCRP, <https://hotcrp.com/>

⁷ [citation needed]

⁸ D. Kaufer and S. Ishizaki, “Computer-aided rhetorical analysis,” in *Applied Natural Language Processing and Content Analysis*, P. M. McCarthy and C. Boonthum, Eds. Hershey, PA: Idea Group Inc., 2011, pp. 275-296.

individual texts to find instances of these types of language to manually compare (the reader will please note that we chose to ignore some considerations of adequate sample size and standard deviations, which are both inconvenient and outside of the scope of this paper). We also calculated average number of words and words per sentence for each review. With this language-level knowledge, we were able to statistically determine characteristics of the best and worst reviewers.

Results

Somewhat contrary to our initial assumptions, our data demonstrate that reviewer B is actually the most gracious and kind of reviewers, while reviewer C is, demonstrably, “the worst.” In sum, our data demonstrate that reviewer C is overall more negative in tone, more likely to dismiss your research question and substitute their own, and on average just longer-winded than reviewer B. We explain these findings below.

Reviewer C has a negative tone

We find that on average, texts from reviewer C have 73% less positive language than texts from reviewer B. Below, we include some excerpts from reviewer B, underlining specific linguistic realizations of their positivity and supportive, uplifting comments.

Important problem, novel ideas, impressive results

The paper is beautifully written

Very nice!

Here we see the reviewer praising the framing of the problem, results, writing, and overall quality of the paper.

In contrast, reviewer C’s comments are much more negative. Below, we include excerpts from reviewer C, again underlining specific linguistic realizations of their negativity (note that some of these comments are regarding the same paper that reviewer B was commenting on, above).

cumbersome notation

conflicting goals

Sorry to say, but...

The results from this section support our overall claim that reviewer C is the least desirable of reviewers, by demonstrating their distinctly bad attitude. In the following sections, we demonstrate two specific tactics that reviewer C uses to display their dissatisfaction with you.

Reviewer C likes to dismiss your research problem and substitute their own

We find that reviewer C takes regular opportunities to explain why the problem you have chosen to explore in this paper is less worthwhile than another research problem (potentially the one he/she studies). See the two examples below.

the reason why a programmer wishes to [use the type of annotations explored in the paper] is to [obtain difficult-to-formalize end guarantees] and not [property you spent weeks proving]

the hard parts of the proof... are in the [part every other paper in the field focuses on, which you explicitly explain in the introduction why you're not]

This tactic not only gives the reviewer a chance to demonstrate their own knowledge and perhaps recommend to you their own work for reference, but also ensures that this review is the most discouraging and hard to rebut because deep down, you know that maybe it's right and you *have* spent the last year answering the wrong question.

Reviewer C is just long-winded

As another tactic for demonstrating their dissatisfaction with you, we find that reviewer C makes their reviews overall more difficult to read and understand. On average, reviews from reviewer C contain 1.5 more words per sentence than those of the kind reviewer B. Additionally, reviews from reviewer C are almost twice as long as those from reviewer B (660 words and 355 words on average, respectively).

We believe that this tactic is meant to ensure not only that you do not finish reading the review, but also that you become sufficiently lost in their long sentences that you cannot fully understand the points they are trying to make.

Conclusion

This study, which was absolutely carefully designed and conducted over a much longer period of time than the day before the SIGBOVIK submission deadline, indicates that authors submitting to conferences with lettered reviews should fear the more negative, longer-winded Reviewer C over the actually pretty reasonable Reviewer B. We expect that this work will lead to the fruitful development of further academic memes. In future work, we hope to gather enough data to support claims about Reviewers A and D, such as how Reviewer A enjoyed your paper but is too junior to fight with anyone at the PC meeting and how Reviewer D took unusual interest in picking apart the formatting of your bibliography.