

Additional Notes

Original paper

Yang Gao, Steffen Eger, Ilya Kuznetsov, Iryna Gurevych, and Yusuke Miyao. "Does My Rebuttal Matter? Insights from a Major NLP Conference." In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 1274-1290. 2019.

RPB in Figure 2

Yuyal Pinter found that the abbreviation "RPB" in Figure 2 was not defined. RPB stands for replicability. It is an aspect to rate in the review template of ACL 2018.

Peer pressure

Nihar Shah pointed out a problem with the term "peer pressure" and aspects of correlation/causality in our paper, which indeed allow room for interpretation and might be misleading to some degree. We make the following clarifications.

The paper uses the term "peer pressure" to refer to the incentive to update the scores and reach a consensus. Such consensus is often explicitly encouraged by the area chairs, e.g. by enforcing a discussion for submissions with high deviation in scores while leaving submissions with low deviation intact. In our analysis, we use the data from the ACL 2018 conference to estimate which factor is more responsible for the score updates, the authors' rebuttal or the opinions of the other reviewers, expressed numerically as scores. For this, we conduct regression analyses. If the scores given by the other reviewers dominate in the regression, we take this as evidence for the peer pressure effect; if factors relating to the rebuttal dominate, we assume the peer pressure effect to be weaker. Our experiments in Section 4.3 show that peer pressure dominates the regression.

Note that both the texts and the scores of the peer reviews can yield peer pressure. A reviewer can be influenced by the ideas expressed in the other reviewers' texts, or simply by the pressure to be aligned with the other reviewers in terms of scores. The latter kind of influence is also known as *herd behavior* (Banerjee, 1992) or *conformity bias* (Buechelet al., 2015). In our study, we do not distinguish the influence of these two factors, but instead treat the influence from the other peer reviews as a whole as "peer pressure". Completely disentangling the causes of peer pressure would require controlled experimentation, e.g. allowing a group of reviewers to only see the scores while allowing another group to only see the comments, and analyze the score-update behavior differences in each group. Such experiments are to be set up in advance and integrated into the reviewing process itself and hence are beyond the scope of our study which operates on historical data. We encourage the community to perform such controlled experiments in the future.

References

Abhijit V. Banerjee. "A simple model of herd behavior." *The Quarterly Journal of Economics* 107, no. 3 (1992): 797-817.

Berno Buechel, Tim Hellmann, and Stefan Klößner. "Opinion dynamics and wisdom under conformity." *Journal of Economic Dynamics and Control* 52 (2015): 240-257.