

Report on EMNLP Reviewer Survey

Aurélie Névéol, Karën Fort, Rebecca Hwa

► **To cite this version:**

Aurélie Névéol, Karën Fort, Rebecca Hwa. Report on EMNLP Reviewer Survey. [Technical Report] Association for computational linguistics. 2017. hal-01660886

HAL Id: hal-01660886

<https://hal.archives-ouvertes.fr/hal-01660886>

Submitted on 11 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Report on EMNLP Reviewer Survey

Aurélie Névéol
Karën Fort
Rebecca Hwa

July 2017

Executive Summary

The EMNLP reviewer survey was conducted in May and June 2017 during the author response and final decision period of the EMNLP 2017 conference. It received 485 complete responses (701 responses in total). This section gives a high-level view of the main results and trends, based on the detailed results presented in subsequent sections.

The sample is representative of the total of 980 reviewers who contributed to the EMNLP program committee. The survey was conducted after the initial EMNLP paper review period (reviewers were not informed of this upcoming survey while reviewing) and before the acceptance decisions were announced (the names of the authors of accepted papers had not been published yet).

The first set of questions directly addressed the robustness of author anonymity during EMNLP 2017 reviewing. Results show that about 30% of survey respondents claim they can identify authors of a paper they reviewed. However, only 19% actually volunteered to submit at least one specific guess of either author or institute that they identified in their EMNLP reviews, suggesting that there may have been some confusion between the general feeling that reviewers could identify authors and the practical exercise of formulating a specific guess relevant to a particular reviewing experience. For instance, 2% of respondents supplied comments instead of guesses, e.g. "I did not try [to identify authors], I simply thought I could do it if I tried".

A total of 122 respondents (17%) supplied 196 unique guesses, 81 of which were completely incorrect, 33 were institution matches, 73 were partial matches (i.e., the guess identified one or more, but not all the authors) and 9 cases were exact matches (i.e., the guess identified all and only the authors).

This suggests that identifying authors is not as easy as it may seem, and that double-blind remains quite blind.

The most frequent reason reported for author identification was that the paper presents work using material or methods specific to the authors (60 %) followed by the fact that the paper addresses a niche topic that only so many colleagues work on (37 %). Having seen the paper on ArXiv is only the third most frequent reason (22 %) given for author identification.

The second set of questions sought to assess reviewers' awareness of potential bias in reviewing practices, as well as reviewers' preferences with respect to the anonymity of their reviews. The large majority of reviewers (71 %) do not wish to sign their reviews, however, only 14 % report that they would decline joining a program committee if signing reviews was mandatory. About 40 % of respondents deny possible influence of knowing the names of the authors or having their name revealed to the authors on the content of their review.

Finally, an open comment field solicited any additional thoughts respondents wished to share. About 22 % of respondents seized this opportunity. A large number of comments discussed reviewing preferences and the practice of preprint publishing, likely in relation with the concomitant ACL survey. We manually identified whether a reviewing preference was expressed, and what that reviewing preference was (with inter-annotator agreement on the analysis of 78%). We found that over 60% of respondents who discussed reviewing preferences are in favor of double-blind reviewing while over 15% discuss reviewing methods without taking a firm stand. We also noted the following recurring themes:

- many respondents are eager for ACL to rule on several aspects of arXiv usage: double submissions to ACL/arXiv (respondents seem in favor of a ban), access of ACL reviewers to arXiv (respondents also seem in favor of a ban), availability of an arXiv anonymous submission feature (respondents suggest ACL should lobby for such a policy as an alternative to double submission ban)
- some respondents are concerned with review quality and would like a review evaluation mechanism involving authors
- some respondents like to have shared reviews between the reviewers of a paper
- there are many reservations with the idea of signing reviews, including the introduction of bias in the reviews and fear of retribution for negative reviews. However some respondents have proposals for mitigating reviewing policies. One original example is

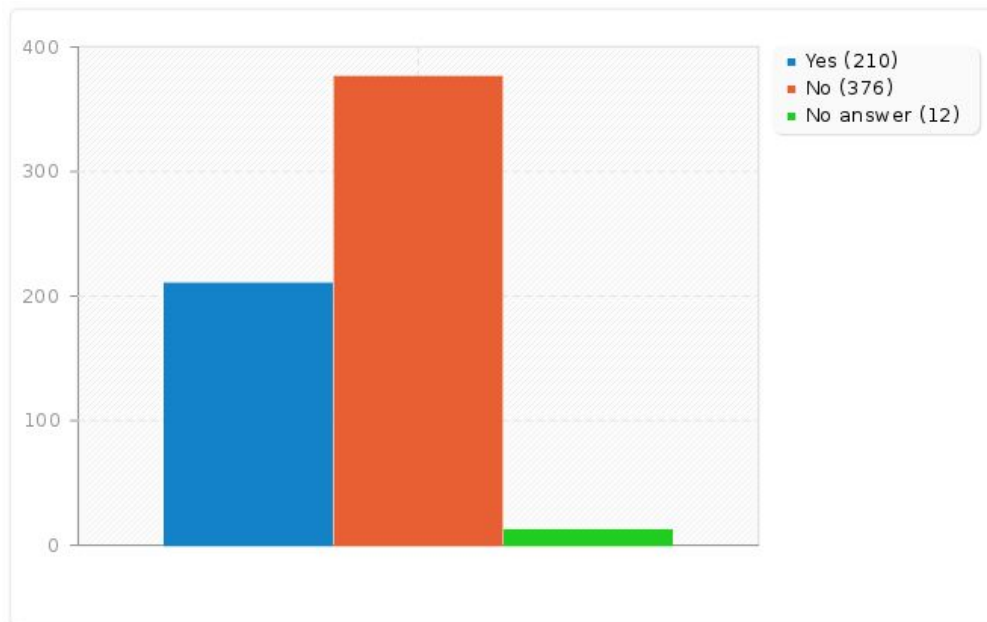
the idea of showing authors the name of reviewers along with the reviews without pairing them.

1. Author Anonymity

In this section, we present the results of the survey questions addressing the robustness of author anonymity. The results are presented based on the total number of responses received. The specific comments quoted may have been modified to preserve confidentiality.

Do you think you can identify at least one author (or institution) from the articles that you reviewed?

While a fair number of respondents felt that could identify at least one author or institution from the articles assigned to them for review, the majority of respondents could not.



Can you supply the alleged name of the author or institution, together with the relevant submission number?

When asked to supply specific guesses related to their reviewing experience at EMNLP 2017, the number of respondents who felt they could supply a guess fell from 210 to 122. This drop in participation between the two questions can be interpreted as a lack of specification of the context of the first question, which some respondents understood as « in any conference that you have previously reviewed for » instead of « EMNLP 2017 ». This is evidenced by some of the comments left by 2% of respondents instead of guesses, e.g.

[This year, I could not do that with confidence as I reviewed papers addressing a wide variety of topics compared to last year where I reviewed papers about a focused topic]

Another reason for the discrepancy is that thinking you can identify authors is different from actually guessing the authors' identity, with or without effort. Some comments also illustrated this, e.g.

[I did not try to identify authors, I simply thought I could do it if I tried].

Some of the respondents also offered general comments or insight on their attitude towards author identification. For example, one respondent deliberately attempted **not** to identify authors:

[I reviewed a paper that was from one of the participants of a workshop at some conference. I could identify them; however, I choose not to look up their name.]

Some reviewers offered insight into their thought process for the specific guesses provided (or not). For instance:

[This submission is basically an XYZ project report; it is trivial to see who runs it. A quick web search identifies Author 1 and Author 2. To clarify: I can guess the authors **now**, but I didn't do the web search before reviewing. Not too many people still work on « Topic T », but I wouldn't have guessed their specific names.] (Interestingly, in spite of the reviewer's confidence in their web search results, neither Author 1 nor Author 2 were authors of the paper commented here)

A total of 122 respondents (17%) supplied 196 unique guesses, 81 of which were completely incorrect, 33 were institution matches, 73 were partial matches (i.e., the guess identified one or more, but not all the authors) and 9 cases were exact matches (i.e., the guess identified all and only the authors). This suggests that identifying authors is not as easy as it may seem, and that double-blind reviewing remains quite blind in the international ACL community. As an indicative comparison, it can be noted that the analysis of guesses supplied by reviewers in the French NLP community who participated in a similar survey based on their reviewing experience for the TALN conference this year revealed that 80% of the guesses were partial or exact matches and an additional 10% were institutional matches. Overall, 50% of the survey respondents had claimed they could identify author information for a paper assigned to them. The French community is much smaller than the ACL community at large, so cases of niche topic or material and methods specific to the authors are more common. However, it can also be noted that one respondent to the French survey reported only submitting guesses that they were certain had to be correct.

Out of the general 196 guesses, participants guessed completely incorrectly 81 times (41%). In the total 196 guesses, 67 were instances of accepted papers (including duplicates); out of these, participants guessed incorrectly 24 times (36%). This suggests the accepted papers might be slightly easier to guess.

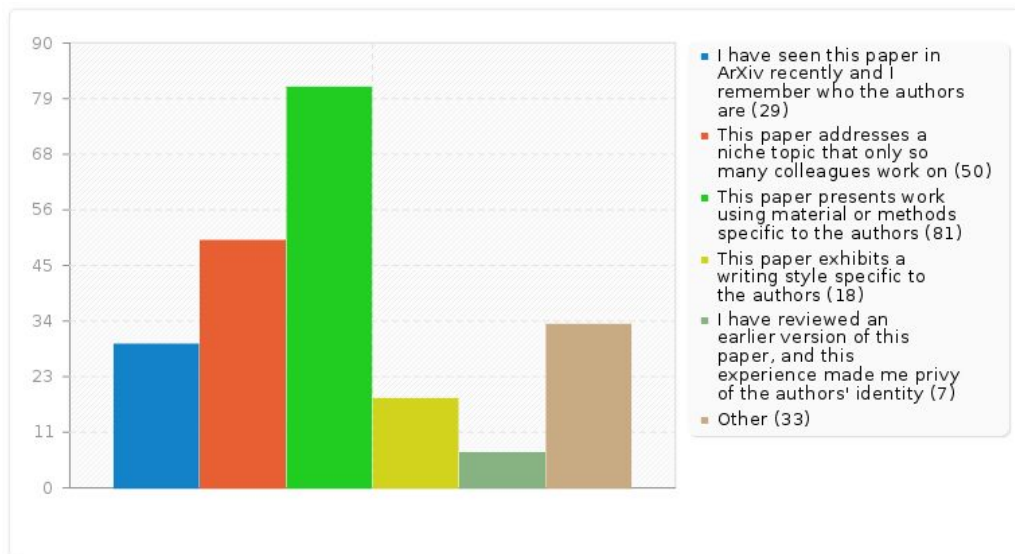
The survey participants guessed on 149 unique papers; the percentage of accepted papers here is 30.2%, which is much higher than the general acceptance rate (22.0%) -- perhaps unsurprisingly, the guesses are biased toward the better papers.

What clues supported your guess of authorship identification?

The most frequent reason reported for author identification was that the paper presents work using material or methods specific to the authors (60 %) followed by the fact that the paper addresses a niche topic that only so many colleagues work on (37 %). Having seen the paper on ArXiv is only the third most frequent reason (22 %) given for author identification.

Most of the « other » reasons for identification reported concerned:

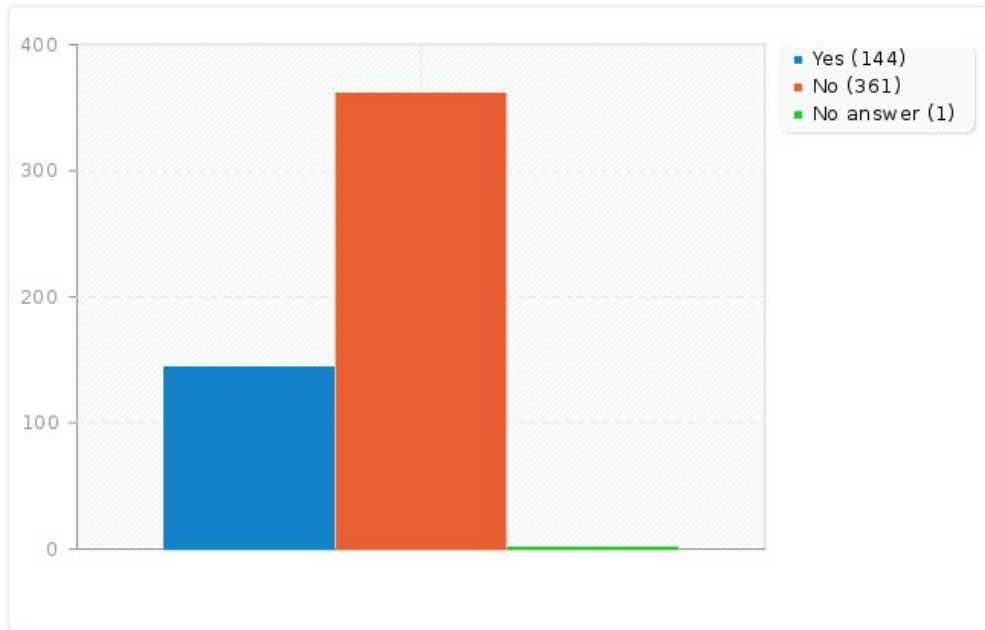
- identifying information left in the submission material such as acknowledgments, supplementary data or code
- content of the related work section that was either felt to re-use content (citation list and/or text) from previous papers of the authors, extensive self-citations, explicit reference to previous work or earlier stages of the work



2. Awareness of bias and reviewing preferences

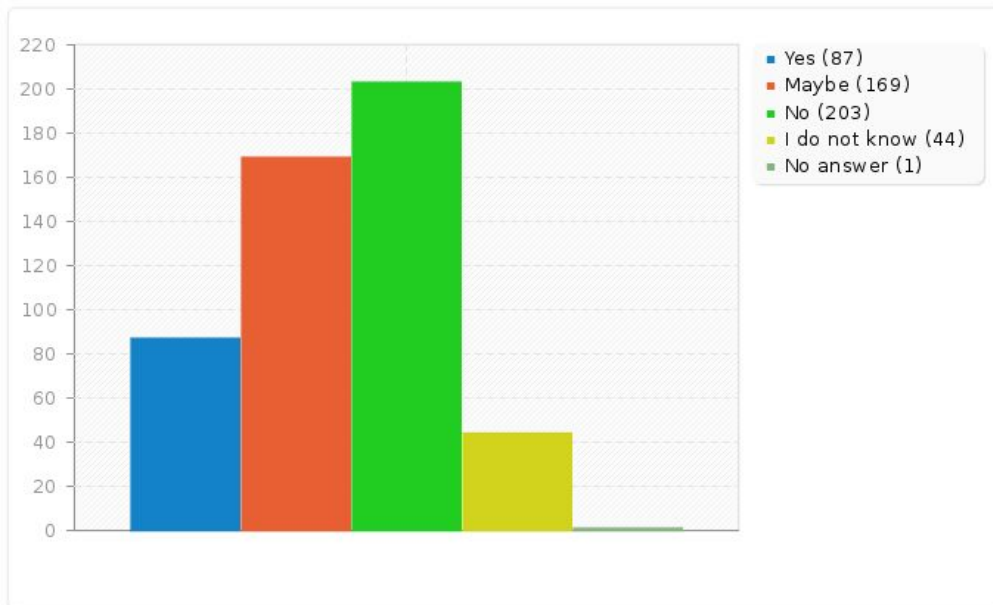
If you had the opportunity to sign your reviews, would you chose to do so?

The large majority of reviewers (71 %) do not wish to sign their reviews.



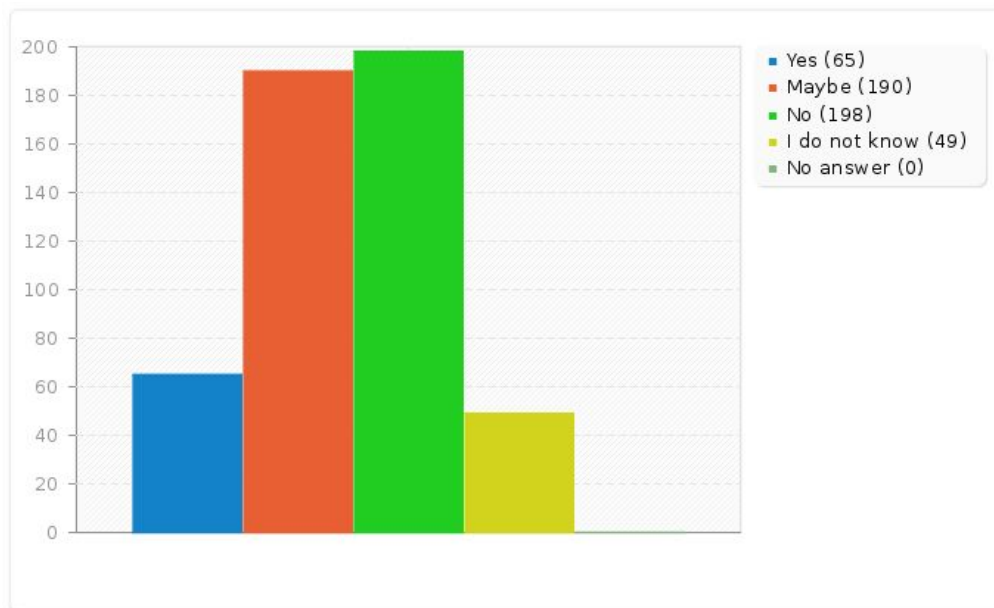
Do you think the content of your reviews would have been different if you had had to sign them?

About 40 % of respondents deny possible bias in their review content due to having their name revealed to the authors. Another 17 % of respondents believe signing reviews will create bias while 34 % think bias is possible, and 9 % do not know what to think.



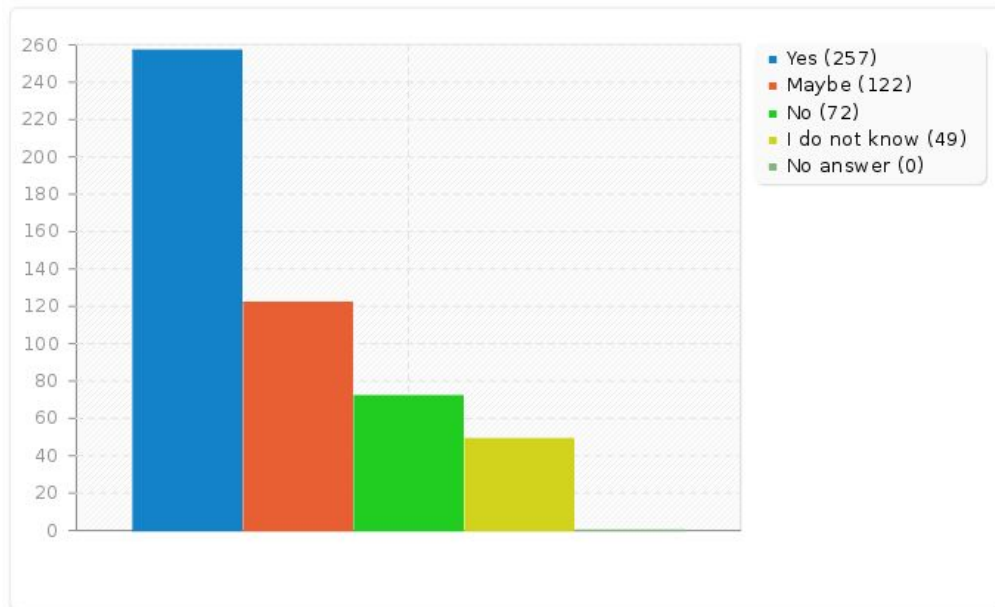
Do you think the content of your reviews would have been different if you had known the name of the authors?

The trend is similar to that observed on the previous question. About 39 % of respondents deny possible bias in their review content due to knowing the name of the authors. Another 13 % of respondents believe signing reviews will create bias while 38 % think bias is possible, and 10 % do not know what to think.



Would you have accepted to be part of the program committee (i.e. act as a reviewer) if you had had to commit to signing your reviews?

Only 14 % of respondents report that they would decline joining a program committee if signing reviews was mandatory. We observe a significant drop between the number of respondents who declare not wanting to sign their reviews (361 respondents answered « no » to the first question in this section) compared to the respondents who declare that having to sign review would lead them to decline participation to a program committee (only 72 respondents answered « no » to this question).



Some comments in the open comment section can perhaps shed some light on this situation. Many respondents reported being unfavorable to signing reviews because they feared negative repercussions in case authors did not like the content of the review. These respondents often reported being early-career researchers and also reported that they felt they could not afford to pass on a reviewing opportunity, so they would feel pressured into accepting to sign reviews if this became a policy. Other respondents reported being established researchers and acknowledged that their status allowed them to not fear repercussions, but they felt that the situation could be different for other members of the community.

3. Open comments

The survey ended with a completely open question asking for additional comments about the survey and the respondents EMNLP reviewing experience. This question was optional and was answered by 153 of the 701 respondents.

The responses encompass comments on the EMNLP reviewing process, on general reviewing practices, opinions and suggestions for future policies regarding reviewing method, assessment of review quality, positioning with respect to preprint publishing. We manually identified whether a reviewing preference was expressed, and what that reviewing preference was (with inter-annotator agreement on the analysis of 78%). Two third of the comments addressed some aspect of reviewing, with 60 % expressing a reviewing preference. Of those, 60% are in favor of double-blind reviewing while over 15% discuss reviewing methods without taking a firm stand. It can be noted that many of the respondents in favor of blind reviewing mentioned bias and felt

that being blind to the author or reviewer identity was an appropriate mitigating strategy. Comments suggest that some respondents are more convinced of the need for reviewer anonymity vs. author anonymity, which translates into favoring single blind reviewing as author anonymity is perceived as unnecessary.

Some respondents felt that the survey did not explore the notion of « signed reviews » quite enough. The point was made that reviewers are currently free to sign their reviews if desired – and yet, few if any, exercise that option. Furthermore, it was pointed out that signed reviews can be handled in a number of ways: the name of the reviewer could be revealed to the authors only, or to the general public if the reviews were published along with a paper. This leads to a notion of imbalance between accepted and non-accepted papers, since the reviews of accepted papers would become more prominent than the reviews of non-accepted papers. In addition, there could be a choice in the timing of reviewer identity reveal, which could occur as soon as reviews were sent to the authors, or upon publication in case of acceptance only. One respondent reported that a review signing policy would lead them to decline reviewing papers when they did not want to support acceptance.

Below is a summary of the main views expressed in the comments, beyond reviewing preference:

- many respondents are eager for ACL to rule on several aspects of arXiv usage: double submissions to ACL/arXiv (respondents seem in favor of a ban), access of ACL reviewers to arXiv (respondents also seem in favor of a ban), availability of an arXiv anonymous submission feature (respondent suggest ACL should lobby for such a policy as an alternative to double submission ban)
- some respondents are concerned with review quality and would like a review evaluation mechanism involving authors
- some respondents like to have shared reviews between the reviewers of a paper
- there are many reservations with the idea of signing reviews, including the introduction of bias in the reviews and fear of retribution for negative reviews. However some respondents have proposals for mitigating reviewing policies. One original example is the idea of showing authors the name of reviewers along with the reviews without pairing them.

Acknowledgements

We would like to acknowledge the EMNLP conference chairs for their support in conducting this survey.