 doi:10.20944/preprints201905.0098.v2

Article

Open Up: A Survey on Open and Non-anonymized Peer Reviewing

Lonni Besançon [1,2]*, Niklas Rönnberg[1], Jonas Löwgren [1], Jonathan P. Tennant [3], and Matthew Cooper [1]

[1] Linköping University, Norrköping, Sweden.
[2] Université Paris Sud, Paris, France.
[3] Institute for Globally Distributed Open Research and Education (IGDORE); University of Southern Denmark; Center for Research and Interdisciplinarity, France.
* Correspondence: lonni.besancon@gmail.com

Abstract: We present a discussion and analysis regarding the benefits and limitations of open and non-anonymized peer review. This analysis is based on literature results and responses to a survey on the reviewing process of alt.chi, a more or less open-review track within the CHI conference, the predominant conference in the field of human-computer interaction (HCI). This track currently is the only implementation of an open-peer-review process in the field of HCI while, with the recent increase in interest in open science practices, open review is now being considered and used in other fields. We collected 30 responses from alt.chi authors and reviewers and found that, while the benefits are quite clear and the system is generally well liked by alt.chi participants, they are reluctant to see it used in other venues. This concurs with a number of recent studies that suggest a divergence between support for a more open review process and its practical implementation. The data and scripts are available on osf.io/vuw7h/, and the figures and follow-up work on the project page.

Keywords: Open Review; Open Science; Zero-Blind Review; Peer Review; Methodology

1. Introduction

Pre-publication peer review of scientific articles is generally considered to be an essential part of ensuring the quality of scholarly research communications [1–3]. It can take many forms from single-round peer-review, typical of conferences, to multiple-stage peer reviewing, more common in scholarly journals. Variants of these processes also include zero-blind (neither reviewers nor authors are anonymous), single-blind (reviewers are anonymous), and double-blind (both authors and reviewers are anonymous) systems (see for example [4]).

The names of these different variations can be confounding. While "open review" has been used in the past to mean 'non-anonymized' reviews (e.g., [5,6]), we will use in this submission "open review" to refer to anonymous or signed reviews that are publicly available. Classical single/double-blind reviewing is held in high regard within scientific communities and is often considered the gold standard for assessing the validity of research communications [1–3,7–10]. Despite the criticism it sometimes incurs [11–17], peer review is still considered to be the "best that we have" [17] and few broad-scale attempts have been made to address the numerous issues with the current system, especially in human-computer interaction (HCI).

The alt.chi conference track, however, is an exception. It is a track within the annual CHI conference, which is the predominant conference in the field of HCI. It started by offering papers rejected from the main track of CHI a second chance to be accepted through a set of different reviewers. The system then evolved into an open (publicly available) and non-anonymous process based on

 doi:10.20944/preprints201905.0098.v2

voluntary reviews. In 2013 and 2018 this approach was changed to a juried process where a small number of reviewers discussed the submissions, but in 2014 and for 2019 reverted to the open, volunteer-based and non-anonymous system.

In this article, we aim to determine what advantages and limitations are presented by open peer-reviewing through both a literature analysis and by gathering opinions from previous alt.chi authors as to what they value from such a system in comparison with the traditional single/double-blind review process. This offers a unique chance to explore an interesting system of peer review, to contribute to our developing understanding of this critical element of scholarly communication.

2. Related Work

Of particular relevance to this discussion is past work on the topic of blind reviews, the benefits and challenges presented by open reviews, and the alternatives adopted in other fields.

2.1. Concerns with peer reviewing

While being almost as old as scholarship itself [18–20], peer-review was only slowly introduced and established as the norm across the scholarly literature. In fact, one anecdote describes how Einstein chose to publish his paper in an alternative journal as an angry reaction to an anonymous peer review, and this may have been Einstein's only actual encounter with peer review [18,21]. While it is now well established, peer review has often been vocally criticized. Recent concerns include but are not limited to (for more, see e.g., [17] or [22]) the lack of adequate training of reviewers, leading to them being unable to detect even major methodological errors [23]; the overall duration of the reviewing process which slows progress in the scientific community [24,25]; the unreliability of the assessments made by reviewers [26,27]; the fact that interesting discussions and mitigation points highlighted by the review process are often not made accessible to other researchers [22]; that the review process is unable to prevent malicious or indifferent reviewers [13]; and that reviewers rarely receive proper credit for their reviews [22]. Noteworthy previous work has concluded that reviewers typically agree on a submitted manuscript at levels only slightly above chance [26] and that the current system of having two or three reviewers is unlikely to do much better than a lottery, based on mathematical modelling [28].

With respect to the CHI conference, Jansen et al. [29] conducted a survey of 46 CHI authors in 2016 to determine what they value in their reviews. Jansen et al. noted that authors appreciated encouragement and having their work fairly assessed but, at the same time, highlighted that authors sometimes found reviews unreasonable or insufficiently detailed. They also discussed and presented several points not covered by the reviewing guidelines (e.g., transparency about the statistical methods used or recommended and why) as well as several methods to make sure these guidelines for reviewers are followed during the reviewing process. They finally argued that the fact that reviews are not public makes it hard to gather data to evaluate the peer review process. This could impede the development of Early Career Researchers (ERCs) who cannot find good examples of reviews from which to learn. These findings were echoed by Squazzoni et al. [30] who argued that the sharing of review data could both encourage and help reward reviewers.

2.2. How blind should it be? The benefits of double-blind reviews

Previous work has already investigated and attempted to summarize the main arguments for and against blinding, reciprocal or not, during peer review [5,31,32]. The four available and most commonly investigated options are zero-blind, single-blind, double-blind, and triple-blind.

Double-blind reviews have been shown by past research to be generally better than single-blind reviews [33–37]. It is thought to reduce reviewers' biases [34,35,37], to increase the number of accepted papers with female first authors in ecology or evolution journals [33], and seems to be generally preferred by both authors and reviewers [36]. Baccheli and Beller [38] showed that, despite the inherent costs of double-blind reviewing (e.g., difficulty for authors to blind papers and difficulty for

 doi:10.20944/preprints201905.0098.v2

reviewers to judge how incremental the work is), less than a third of the surveyed software engineering community disagreed with a switch from single-blind reviewing to double-blind reviewing. Prechelt et al. [15] investigated the perception of peer reviewing in the same community and reported that only a third of reviews are considered useful while the rest are seen as unhelpful or misleading. Many respondents to their survey supported the adoption of either double-blind or zero-blind reviewing.

With respect to the effectiveness of anonymizing authors there is conflicting evidence [39]. Part of the literature argues that hiding their identity leads to better and less biased reviews [40–42], while it would seem that several large scale studies do not support such claims [43–46]. Still, anonymizing authors appears to be one of the best solutions to address the known bias in research communities against female scientists and to increase the overall diversity of researchers [47–49].

Double-blind reviewing cannot, however, solve all the concerns previously mentioned but open peer review might yield interesting answers to these concerns.

2.3. Towards (Anonymous) Open Peer Review

With the recent publicity surrounding open research and open access, it might seem that open peer reviewing is new. However, journals practising open reviews have existed since at least the 1990s [50] and the possible benefits of open peer reviews have been widely discussed in the literature (e.g., [51]). The sharing of review reports in one form or another actually even goes back to the origins of peer review itself [52]. The term "open review" is, however, loosely used and encompasses several elements [17,53] that should be distinguished [54]: open identities, open reports, open participation, open interaction, open pre-review manuscripts, open final-version commenting and open platforms. As stated in the introduction, in this manuscript we wish to at least distinguish between openly available reviews and non-anonymized peer reviews. We feel that the best way for open peer review to progress is for different communities to advance the different elements outlined above, based on the best evidence available to them about what works best.

Jones [55] argued that anonymization could be detrimental because reviewers could act without fear of sanctions and suggested that reviews should be signed. This was also supported by Shapiro [56]. There are many variations on anonymity [22]. For example the identities of reviewers could be revealed only on published papers while reviewers of rejected papers maintain their anonymity (as is current practice in Frontiers in Neuroscience [57]), or reviewers could have to directly sign their reviews. Similarly, one has to distinguish between revealing the reviewers' identities only to the authors or to the public by adding the names of the reviewers to the published manuscript, often (though not always) accompanied by their report and interactions with the authors. PeerJ gives the reviewers the option to add their names to their reports and the authors the possibility to add all interactions made during the reviewing process to the published manuscript [58] while BMC Public Health (and other BMC series) have made publication of signed reviews standard practice [59]. Yet another form of openness is to publish unsigned reviewers' reports (which we define as open, anonymous peer review). This system is currently used by, for example, The American Journal of Bioethics [60].

The benefits of an open and/or non-anonymized reviewing system have been identified or postulated in previous work. Based on their investigation of peer-review based learning to foster learning of students with heterogeneous backgrounds, Pucker et al. [61] expected that "Reviewers might be more motivated thus producing better reports when they know that their reports will be published. In addition, errors in reviews could be identified and removed if a large number of peers are inspecting them." Signed reviews have been evaluated as more polite and of higher quality when compared to anonymous reviews even though the reviewing process was found to be longer [51,62].

3. Polling the alt.chi community on open review

Within HCI we know of only one forum that uses an open-review process: the alt.chi track within the CHI conference. Its initial purpose was to offer rejected papers a second chance through another round of peer-reviewing with new reviewers. Over the years, it has changed many times to include an

 doi:10.20944/preprints201905.0098.v2

open and public reviewing process or, in some years, a juried process. The procedure for open and public reviewing with open-participation is the following:

- Authors submit a non-anonymized manuscript to a public forum.
- Anyone can submit a review or discuss the paper. Authors can invite reviewers.
- To ensure a sufficient number of reviews, authors of submissions are asked to review other submissions.
- Reviews are published non-anonymously. Anyone, authors and other reviewers, can see and respond to them until the system closes.
- The system is closed and some submissions are accepted and presented at the conference. In some cases, authors are asked to attach the reviews and discussions to their manuscript.

To better understand the advantages and limitations of such a review process in the HCI community, we asked previous authors to complete a short survey on the reviewing system that was in place at alt.chi. We first gathered the contact information of at least the first author of every alt.chi paper from 2010 to 2018. When we believed that the first author of a publication could have already been the first author of an other publication, we also added the last author contact email to our list. We then sent an email to all identified contacts providing a link to the survey. Additionally, we repeatedly posted a link on Twitter with the hashtag 'chi2019', and asked people to forward the survey as much as possible. The online survey is still available, though closed to new responses.

The survey comprised different categories of questions. The first category was about the person's point of view as an author (Sect. B). The second explored the person's point of view as an alt.chi reviewer (Sect. C). A final category (Sect. D) evaluated how each respondent felt about the reviewing process and whether they would continue using it within alt.chi and even extend it to other tracks. We also sought to gather additional comments about peer review and the questionnaire itself.

4. Results and Discussion

We gathered a total of 30 responses to our survey. We initially had 31 responses but one respondent did not confirm that we could use their answers in a future publication so we removed their response from our data. While such a low number of respondents could be potentially seen as problematic, it appears through the literature that, to gather subjective measures and opinions, it can be enough. Indeed, Isenberg et al. [63] showed that on average between 1—5 participants are used in evaluation of research projects, while Caine [64] showed that among all CHI papers published in one year, all of the papers comprising user studies and therefore reporting on qualitative feedback and/or quantitative measures had less than 30 respondents/participants on average. Similar findings were reported in a more recent look at studies and participants [65]: in interviews or lab studies (both of which contain qualitative feedback and/or quantitative Likert-scale ratings) the majority of studies are conducted with fewer than 20 participants. In fact, for qualitative feedback and quantitative answers to Likert-scale the average is likely to be even lower and we found that often such research projects report results with 15 or less respondents (e.g., [66–72]), and sometimes with numbers as low as one (e.g., [68]) or two (e.g., [69]). Finally, we argue based on the literature, that there is no meaningful cut-off point at which a sample size becomes inadequate or invalid because it would be "too small" [73] but instead the relationship between the value of a study and the size of the sample incrementally increases with each additional participant [73].

All anonymized answers (quantitative and qualitative) and scripts used on the data are available at https://osf.io/vuw7h/. Respondents had submitted an average of 1.9 papers (SD = 1.8) through the open reviewing process of alt.chi, while only two authors had submitted to a juried version of alt.chi. Most respondents (26) had submitted more than ten papers to more classical review tracks and were experienced with single/double blind reviewing. The other four respondents had submitted between one and ten papers to other venues. Respondents had reviewed an average of 8.4 papers for alt.chi (SD = 10.1), while only three of them had reviewed for the juried process of alt.chi 2018. 26 respondents

 doi:10.20944/preprints201905.0098.v2

had reviewed more than ten papers in a single/double blind review process while the remaining four had reviewed between one and ten papers within such a process.

4.1. Qualitative feedback: limitations and advantages of the alt.chi reviewing process

To limit interpretation biases when analysing the answers to open-ended questions, one of the five investigators did a first pass to categorize each comment. Two other investigators used these categories to classify comments. We consider that an answer belongs to a category if two or more of the three authors classified it as belonging to that category. Our categorization spreadsheet is also available at https://osf.io/vuw7h/.

Concerning the alt.chi process in particular, respondents highlighted the limits of the invite-to-review (i.e., open participation) system as authors could invite friends to review (2 respondents), papers deemed uninteresting had less chance of acceptance as they would receive less reviews (4 respondents), or the reviewing could simply be a popularity contest in the end making individual reviews less relevant (7 respondents).

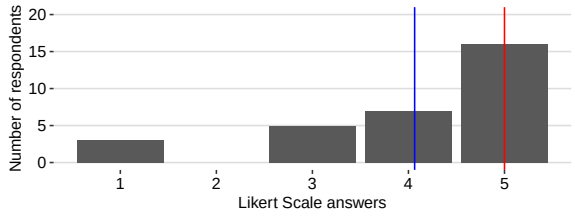
Overall, respondents praised the discussions that the open review process of alt.chi brings, which is an advantage for both authors (13 respondents) and reviewers (14 respondents) and can also stimulate the discussions between reviewers (3 respondents). The added transparency in the reviewing process was praised by five respondents as a benefit for authors as it helps them understand the comments from reviewers (2 respondents) and can reduce the *cite-me* effect (1 respondent). They mentioned that reviewers used a more polite tone (4 respondents), mirroring previous literature findings [51,62], that it fosters future collaborations as authors can directly contact reviewers and vice-versa (2 respondents), and that the more diverse set of reviewers could also lead to interesting discussions (2 respondents). Extending Jansen et al.'s [29] findings, respondents highlighted that reviewers' comments are usually better justified because reviewers are directly accountable for their reviews: this was seen as an advantage for both authors (6 respondents) and reviewers (8 respondents). Interestingly, three respondents mentioned that signing reviews was a good way to receive credit for their work.

Reinforcing findings from previous research, some respondents expressed concerns that an author's reputation could possibly directly influence the reviewer and the decision on the submission (4 respondents as a limitation for authors, 2 for reviewers) or the fact that reviewers might fear being truly critical and, consequently, self-censor their reviews (14 respondents). Finally, four respondents mentioned that negative reviews, even if well-founded, could generate animosity and result in retaliation with respect to future submissions by the reviewer. This echoes past studies highlighting that when given the choice, most reviewers would not sign their reports [74].

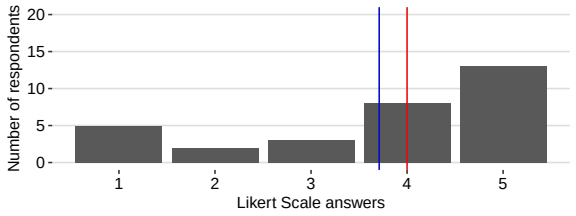
4.2. Quantitative results: would the community consider this process for other CHI tracks?

We have gathered the results of Likert Scale Ratings (questions 11 to 14) in Fig. 1a to Fig. 1d. For all questions a score of 1 indicates 'I disagree' and a score of 5 'I agree'. We present these results with a bar chart showing the ranges of responses (as usually recommended [75]) in addition to means and medians. While the use of means for ordinal values has been initially slightly advocated against [76] and is still highly controversial [77], it appears in the literature that it is nonetheless highly used [78], useful to present [76,77,79,80], and potentially even more useful than medians [79,81]. The results in Fig. 1a and Fig. 1b highlight the openness and interest towards an open and non-anonymous review process that was already suggested by our qualitative results. Indeed, 23 respondents gave a score of 4 or 5 (mean = 4.06, median = 5) to open review and 21 gave a score of 4 or 5 (mean= 3.71, median = 4) to non-anonymous reviews. This is not surprising since respondents have experience with this reviewing process for alt.chi. However, when asked whether they would consider such a process for all other CHI tracks the results diverged from this. It seems that making reviews public (but not anonymous, Fig. 1c) could be envisioned, as 16 respondents would consider it and gave a score of 4 or 5 (mean = 3.29, median = 4). However, concerning the possibility to sign reviews, most

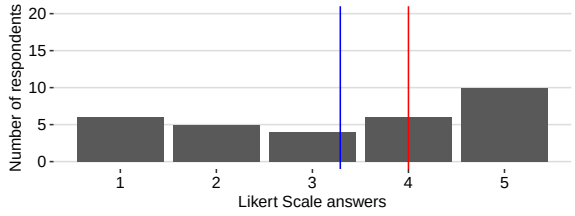
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2019 doi:10.20944/preprints201905.0098.v2



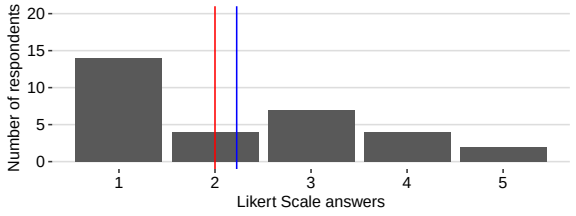
(a) I would consider an open reviewing process for alt.chi.



(b) I would consider a non-anonymous reviewing process for alt.chi.



(c) I would consider an open reviewing process for all CHI submissions.



(d) I would consider a non-anonymous reviewing process for all CHI submissions.

Figure 1. Results of the Likert Scale Ratings for each question that participants were asked. The red bar indicates the median, the blue bar the mean.

respondents would not consider it: 18 gave scores of 1 or 2 (mean = 2.23, median = 2). This mirrors the qualitative feedback regarding the possibility of such a process incurring retaliation for the reviewers of a rejected paper, for example, and echoes previous work (e.g., [9,54]). Several possible procedures for non-anonymous reviews exist beyond simply asking reviewers to sign their reviews, however, such as giving the names of reviewers without attaching them to any specific report or only publishing the names of reviewers of accepted papers. Such alternatives are, however, still rarely used and we hypothetize that they were probably not considered by most of our respondents (though future work should probably investigate this aspect further). Nonetheless, the reluctance to sign reviews for other CHI tracks contrasts with the now quite high number of journals that are using non-anonymous and public reviews (see e.g., some of the BMC series [59] and the transpose-publishing site for a complete list).

While these results are interesting and could potentially help argue for opening the reviewing process to make reviews public, even if not signed, one has to take into account that respondents were all previously involved with alt.chi and should therefore be considered likely to be more open to the process than the rest of the community. It is therefore difficult to guarantee that the rather positive views towards open reviews would be shared by the larger CHI community. A possible follow-up to our work could include gathering all the reviews and discussions generated through an instance of alt.chi and sharing it with the CHI community to produce a more diverse but informed opinion. In any case, future work includes polling authors and reviewers of the CHI community that do not participate in the alt.chi process in order to see if their opinions and ratings diverge from the ones of alt.chi participants.

5. Conclusion

We have conducted an initial investigation on the perception of open-reviewing within the only venue that has an open-reviewing process in the Human-Computer Interaction community. This initial work highlighted that the non-anonymous open reviewing process adopted at alt.chi has some inherent flaws in its open participation design that could easily be addressed while maintaining the overall open (reports) and non-anonymous process. For instance, having a fixed number of assigned reviewers could solve many of the issues mentioned with the alt.chi system. From our results it seems safe to assume that much of the alt.chi community values open and non-anonymous reviewing in

general, but understanding the extent of this will require more work beyond our pilot investigation here. It would also seem that the alt.chi community fears that the implementation of non-anonymous reviews within a more serious venue (a more prestigious venue with a more rigorous review process) could lead to issues; mostly around biases towards accepting the work of more established researchers, self-censorship of reviews, or the possibility for authors to hold a grudge against their reviewers. While other scientific communities are starting to embrace the benefits of open and non-anonymous peer reviewing, the HCI community's only implementation of it remains at alt.chi which only counts as extended abstract rather than a full publication in the proceedings of the conference. Based on our empirical findings, it seems hard to challenge the old belief that 'double-blind peer review is the worst academic QA system, except for all the others'. We conclude and hope that this submission helps to open up a discussion about the fact more open peer-reviewing processes could be tested at alternative venues to alt.chi, with an implementation that avoids the problems identified in this submission. The small-scale survey implemented here could easily be adapted to help other communities understand and optimize their own peer review processes.

Author Contributions:

Lonni Besançon: Conceptualization, Data curation, Investigation, Project administration, Validation, Software, Visualization, Writing - original draft
Niklas Rönnberg: Writing - original draft, Data curation
Jonas Löwgren: Writing - original draft, Writing, review & editing
Jonathan P. Tennant: Writing - review & editing
Matthew Cooper: Writing - original draft, Data curation

Funding: This research received no external funding

Acknowledgments: We wish to thank, the respondents to the survey and Pierre Dragicevic for insightful feedback.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HCI	Human Computer Interaction
CHI	The Computer Human Interaction conference, the main venue for HCI researchers.
alt.chi	The only track within the CHI conference and the whole HCI community that implements an open-review process.

Appendix A.

Appendix B. Questions as an author

1 How many papers have you submitted to alt.chi before CHI2018? (Open)
2 How many papers have you submitted to alt.chi with the juried selection process (i.e., how many papers have you submitted to alt.chi in 2018)? (Open)
3 How many papers have you already submitted to venues with a double/single blind reviewing process (i.e., for which reviewing was anonymous and not open)? (Possible answers: 0, 1–10, 10+)
4 What do you think are the advantages for authors with the open/public and non-anonymized reviewing that was in place before CHI2018 when compared to the traditional double blind reviewing process? (Open)
5 What do you think are the drawbacks/limitations for authors with the open/public and non-anonymized reviewing that was in place before CHI2018 when compared to the traditional double-blind reviewing process? (Open)

Appendix C. Questions as a reviewer

6 How many papers have you reviewed for alt.chi before CHI2018? (Open answer)

7 Have you reviewed for alt.chi in 2018? (Yes or No)
8 How many papers have you reviewed for other venues with a double/single blind reviewing process (i.e., for which reviewing was anonymous and not open)? (Possible answers: 0, 1–10, 10+)
9 What do you think are the advantages for reviewers with the open/public and non-anonymized reviewing that was in place before CHI2018 when compared to the traditional double/single blind reviewing process?
10 What do you think are the drawbacks/limitations for reviewers with the open/public and non-anonymized reviewing that was in place before CHI2018 when compared to the traditional double/single blind reviewing process?

Appendix D. Additional questions

11 I would consider an open/public (but possibly anonymous) reviewing process for all future alt.chi submissions. (Likert scale from 1 to 5 with 1 = "I disagree" and 5 = "I agree")
12 I would consider a non-anonymized reviewing process for all future alt.chi submissions. (Likert scale from 1 to 5 with 1 = "I disagree" and 5 = "I agree")
13 I would consider an open/public (but possibly anonymous) reviewing process for all CHI submissions. (Likert scale from 1 to 5 with 1 = "I disagree" and 5 = "I agree")
14 I would consider a non-anonymized reviewing process for all CHI submissions. (Likert scale from 1 to 5 with 1 = "I disagree" and 5 = "I agree")
15 If you wish to receive the results of our survey, you can enter your e-mail here. This information will not be used when making the data available. (Open Answer)
16 Do you allow us to use the information you provided in future submission (once correctly anonymized)? (Possible answers: Yes or No)
17 Do you have any additional comments on peer review ? (Open answer)
18 Do you have any additional comments on the questionnaire itself?

1. Morgan, P.P. Anonymity in medical journals. *Canadian Medical Association journal* **1984**, *131*, 1007–1008.
2. Pierson, C.A. Peer review and journal quality. *Journal of the American Association of Nurse Practitioners* **2018**, *30*.
3. Wilson, J.D. Peer review and publication. Presidential address before the 70th annual meeting of the American Society for Clinical Investigation, San Francisco, California, 30 April 1978. *The Journal of Clinical Investigation* **1978**, *61*, 1697–1701. doi:10.1172/JCI109091.
4. Largent, E.A.; Snodgrass, R.T., Chapter 5 Blind Peer Review by Academic Journals. In *Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law*; Robertson, C.T.; Kesselheim, A., Eds.; Academic Press, 2016; pp. 75–95. doi:10.1016/b978-0-12-802460-7.00005-x.
5. Pontille, D.; Torny, D. The blind shall see! the question of anonymity in journal peer review. *Ada: A Journal of Gender, New Media, and Technology* **2014**, *4*. doi:10.7264/N3542KV.
6. Ross-Hellauer, T. What is open peer review? A systematic review. *F1000Research* **2017**, *6*. doi:10.12688/f1000research.11369.2.
7. Baggs, J.G.; Broome, M.E.; Dougherty, M.C.; Freda, M.C.; Kearney, M.H. Blinding in peer review: the preferences of reviewers for nursing journals. *Journal of Advanced Nursing* **2008**, *64*, 131–138, [https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2648.2008.04816.x]. doi:10.1111/j.1365-2648.2008.04816.x.
8. Haider, J.; Åström, F. Dimensions of trust in scholarly communication: Problematizing peer review in the aftermath of John Bohannon's "Sting" in science. *Journal of the Association for Information Science and Technology* **2016**, *68*, 450–467, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23669]. doi:10.1002/asi.23669.
9. Mulligan, A.; Hall, L.; Raphael, E. Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, *64*, 132–161, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.22798]. doi:10.1002/asi.22798.

10. Moore, S.; Neylon, C.; Eve, M.P.; O'Donnell, D.P.; Pattinson, D. "Excellence R Us": university research and the fetishisation of excellence. *Palgrave Communications* **2017**, *3*, 16105. doi:dx.doi.org/10.1057/palcomms.2016.105.
11. Armstrong, J.S. Peer review for journals: Evidence on quality control, fairness, and innovation. *Science and Engineering Ethics* **1997**, *3*, 63–84. doi:10.1007/s11948-997-0017-3.
12. Baxt, W.G.; Waeckerle, J.F.; Berlin, J.A.; Callaham, M.L. Who Reviews the Reviewers? Feasibility of Using a Fictitious Manuscript to Evaluate Peer Reviewer Performance. *Annals of Emergency Medicine* **1998**, *32*, 310–317. doi:10.1016/S0196-0644(98)70006-X.
13. D'Andrea, R.; O'Dwyer, J.P. Can editors save peer review from peer reviewers? *PLOS ONE* **2017**, *12*, 1–14. doi:10.1371/journal.pone.0186111.
14. Hettyey, A.; Griggio, M.; Mann, M.; Raveh, S.; Schaedelin, F.C.; Thonhauser, K.E.; Thoß, M.; van Dongen, W.F.D.; White, J.; Zala, S.M.; Penn, D.J. Peerage of Science: will it work? *Trends in Ecology & Evolution* **2012**, *27*, 189–190. doi:10.1016/j.tree.2012.01.005.
15. Prechelt, L.; Graziotin, D.; Fernández, D.M. A community's perspective on the status and future of peer review in software engineering. *Information and Software Technology* **2018**, *95*, 75 – 85. doi:https://doi.org/10.1016/j.infsof.2017.10.019.
16. Tennant, J.; Dugan, J.; Graziotin, D.; Jacques, D.; Waldner, F.; Mietchen, D.; Elkhatib, Y.; B. Collister, L.; Pikas, C.; Crick, T.; Masuzzo, P.; Caravaggi, A.; Berg, D.; Niemeyer, K.; Ross-Hellauer, T.; Mannheimer, S.; Rigling, L.; Katz, D.; Greshake Tzovaras, B.; Pacheco-Mendoza, J.; Fatima, N.; Poblet, M.; Isaakidis, M.; Irawan, D.; Renaut, S.; Madan, C.; Matthias, L.; Nrgaard KjÊr, J.; O'Donnell, D.; Neylon, C.; Kearns, S.; Selvaraju, M.; Colomb, J. A multi-disciplinary perspective on emergent and future innovations in peer review [version 3; referees: 2 approved]. *F1000Research* **2017**, *6*. doi:10.12688/f1000research.12037.3.
17. Tennant, J.P. The state of the art in peer review. *FEMS Microbiology Letters* **2018**, *365*, fny204. doi:10.1093/femsle/fny204.
18. Baldwin, M. In referees we trust? *Physics Today* **2017**, *70*, 44–49. doi:10.1063/pt.3.3463.
19. Baldwin, M. What it was like to be peer reviewed in the 1860s. *Physics Today* **2017**. doi:https://doi.org/10.1063/PT.3.3463.
20. Spier, R. The history of the peer-review process. *Trends in Biotechnology* **2002**, *20*, 357 – 358. doi:https://doi.org/10.1016/S0167-7799(02)01985-6.
21. Kennefick, D. Einstein versus the physical review. *Physics Today* **2005**, *58*, 43. doi:https://doi.org/10.1063/1.2117822.
22. Walker, R.; Rocha da Silva, P. Emerging trends in peer review—a survey. *Frontiers in Neuroscience* **2015**, *9*, 169. doi:10.3389/fnins.2015.00169.
23. Schroter, S.; Black, N.; Evans, S.; Carpenter, J.; Godlee, F.; Smith, R. Effects of training on quality of peer review: randomised controlled trial. *BMJ* **2004**, *328*, 673, [https://www.bmj.com/content/328/7441/673.full.pdf]. doi:10.1136/bmj.38023.700775.AE.
24. Bornmann, L.; Daniel, H.D. How long is the peer review process for journal manuscripts? A case study on Angewandte Chemie International Edition. *CHIMIA International Journal for Chemistry* **2010**, *64*, 72–77. doi:doi:10.2533/chimia.2010.72.
25. Benos, D.J.; Bashari, E.; Chaves, J.M.; Gaggar, A.; Kapoor, N.; LaFrance, M.; Mans, R.; Mayhew, D.; McGowan, S.; Polter, A.; Qadri, Y.; Sarfare, S.; Schultz, K.; Splittgerber, R.; Stephenson, J.; Tower, C.; Walton, R.G.; Zotov, A. The ups and downs of peer review. *Advances in Physiology Education* **2007**, *31*, 145–152, [https://doi.org/10.1152/advan.00104.2006]. PMID: 17562902, doi:10.1152/advan.00104.2006.
26. Kravitz, R.L.; Franks, P.; Feldman, M.D.; Gerrity, M.; Byrne, C.; Tierney, W.M. Editorial Peer Reviewers' Recommendations at a General Medical Journal: Are They Reliable and Do Editors Care? *PLOS ONE* **2010**, *5*, 1–5. doi:10.1371/journal.pone.0010072.
27. Mahoney, M.J. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* **1977**, *1*, 161–175. doi:10.1007/BF01173636.
28. Herron, D.M. Is expert peer review obsolete? A model suggests that post-publication reader review may exceed the accuracy of traditional peer review. *Surgical Endoscopy* **2012**, *26*, 2275–2280. doi:10.1007/s00464-012-2171-1.

29. Jansen, Y.; Hornbaek, K.; Dragicevic, P. What Did Authors Value in the CHI'16 Reviews They Received? Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 2016. doi:10.1145/2851581.2892576.
30. Squazzoni, F.; Grimaldo, F.; Marušić, A. Publishing: Journals could share peer-review data. *Nature* **2017**, *546*, 352.
31. Jubb, M. Peer review: The current landscape and future trends. *Learned Publishing* **2016**, *29*, 13–21, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/leap.1008]. doi:10.1002/leap.1008.
32. Snodgrass, R. Single- Versus Double-blind Reviewing: An Analysis of the Literature. *SIGMOD Rec.* **2006**, *35*, 8–21. doi:10.1145/1168092.1168094.
33. Budden, A.E.; Tregenza, T.; Aarssen, L.W.; Koricheva, J.; Leimu, R.; Lortie, C.J. Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution* **2008**, *23*, 4 – 6. doi:https://doi.org/10.1016/j.tree.2007.07.008.
34. Jefferson, T.; Godlee, F. *Peer review in health sciences*; BMJ Books, 2003.
35. Kassirer, J.; Campion, E. Peer review: Crude and understudied, but indispensable. *JAMA* **1994**, *272*, 96–97. doi:10.1001/jama.1994.03520020022005.
36. Regehr, G.; Bordage, G. To blind or not to blind? What authors and reviewers prefer. *Medical Education* **2006**, *40*, 832–839, [https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2929.2006.02539.x]. doi:10.1111/j.1365-2929.2006.02539.x.
37. Ross, J.S.; Gross, C.P.; Desai, M.M.; Hong, Y.; Grant, A.O.; Daniels, S.R.; Hachinski, V.C.; Gibbons, R.J.; Gardner, T.J.; Krumholz, H.M. Effect of blinded peer review on abstract acceptance. *JAMA* **2006**, *295*, 1675–1680, [/data/journals/jama/5019/joc60040.pdf]. doi:10.1001/jama.295.14.1675.
38. Bacchelli, A.; Beller, M. Double-Blind Review in Software Engineering Venues: The Community's Perspective. 2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C), 2017, pp. 385–396. doi:10.1109/ICSE-C.2017.49.
39. Tennant, J.P. The dark side of peer review. EON, 2017, Vol. 10, pp. 2–4. doi:10.18243/eon/2017.10.8.1.
40. McNutt, R.; Evans, A.; Fletcher, R.; Fletcher, S. The effects of blinding on the quality of peer review: A randomized trial. *JAMA* **1990**, *263*, 1371–1376. doi:10.1001/jama.1990.03440100079012.
41. Baggs, J.G.; Broome, M.E.; Dougherty, M.C.; Freda, M.C.; Kearney, M.H. Blinding in peer review: the preferences of reviewers for nursing journals. *Journal of advanced nursing* **2008**, *64*, 131–138.
42. Weicher, M. Peer review and secrecy in the "Information Age". *Proceedings of the American Society for Information Science and Technology* **2008**, *45*, 1–12.
43. Isenberg, S.J.; Sanchez, E.; Zafran, K.C. The effect of masking manuscripts for the peer-review process of an ophthalmic journal. *British Journal of Ophthalmology* **2009**, *93*, 881–884, [https://bjo.bmj.com/content/93/7/881.full.pdf]. doi:10.1136/bjo.2008.151886.
44. Justice, A.; Cho, M.; Winker, M.; Berlin, J.; Rennie, D.; the PEER Investigators. Does masking author identity improve peer review quality? A randomized controlled trial. *JAMA* **1998**, *280*, 240–242, [/data/journals/jama/4568/jpv71028.pdf]. doi:10.1001/jama.280.3.240.
45. Lee, C.J.; Sugimoto, C.R.; Zhang, G.; Cronin, B. Bias in peer review. *Journal of the American Society for Information Science and Technology* **2013**, *64*, 2–17. doi:10.1002/asi.22784.
46. Van Rooyen, S.; Godlee, F.; Evans, S.; Smith, R.; Black, N. Effect of blinding and unmasking on the quality of peer review: A randomized trial. *JAMA* **1998**, *280*, 234–237, [/data/journals/jama/4568/jpv71017.pdf]. doi:10.1001/jama.280.3.234.
47. Darling, E.S. Use of double-blind peer review to increase author diversity. *Conservation Biology* **2015**, *29*, 297–299. doi:10.1111/cobi.12333.
48. Helmer, M.; Schottdorf, M.; Neef, A.; Battaglia, D. Research: Gender bias in scholarly peer review. *eLife* **2017**, *6*, e21718. doi:10.7554/eLife.21718.
49. Roberts, S.G.; Verhoef, T. Double-blind reviewing at EvoLang 11 reveals gender bias†. *Journal of Language Evolution* **2016**, *1*, 163–167. doi:10.1093/jole/lzw009.
50. Parks S, G.S. *Tracking Global Trends in Open Peer Review*.
51. Walsh, E.; Rooney, M.; Appleby, L.; Wilkinson, G. Open peer review: A randomised controlled trial. *British Journal of Psychiatry* **2000**, *176*, 47–51. doi:10.1192/bjp.176.1.47.
52. Csiszar, A. Peer review: Troubled from the start. *Nature News* **2016**, *532*, 306. doi:10.1038/532306a.

53. Ross-Hellauer, T.; Schmidt, B.; Kramer, B. Are funder Open Access platforms a good idea? *PeerJ Preprints* **2018**, *6*, e26954v1. doi:10.7287/peerj.preprints.26954v1.
54. Ross-Hellauer, T.; Deppe, A.; Schmidt, B. Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers. *PLOS ONE* **2017**, *12*, e0189311–.
55. Jones, R. Rights, wrongs and referees. *New Scientist* **1974**, *61*, 758–759.
56. Shapiro, B.J. *A culture of fact: England, 1550-1720*; Cornell University Press, 2003.
57. in Neuroscience, F. *Frontiers in Neuroscience Review System*.
58. PeerJ. *Policies and Procedures*.
59. Health, B.P. *Peer Review Policy*.
60. of Bioethics, T.A.J. *Standards for Manuscript Submission General Information*.
61. Pucker, B.; Schilbert, H.; Schumacher, S.F. Integrating molecular biology and bioinformatics education. *Preprints 2018* **2018**. doi:10.20944/preprints201811.0183.v1.
62. Snell, L.; Spencer, J. Reviewers' perceptions of the peer review process for a medical education journal. *Medical Education* **2005**, *39*, 90–97, [https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2929.2004.02026.x]. doi:10.1111/j.1365-2929.2004.02026.x.
63. Isenberg, T.; Isenberg, P.; Chen, J.; Sedlmair, M.; Möller, T. A Systematic Review on the Practice of Evaluating Visualization. *IEEE Transactions on Visualization and Computer Graphics* **2013**, *19*, 2818–2827. doi:10.1109/TVCG.2013.126.
64. Caine, K. Local Standards for Sample Size at CHI. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; ACM: New York, NY, USA, 2016; CHI '16, pp. 981–992. doi:10.1145/2858036.2858498.
65. Koeman, L. How many participants do researchers recruit? A look at 678 UX/HCI studies. Online. Last visited 06 January 2019, 2018.
66. Besançon, L.; Semmo, A.; Biau, D.J.; Frachet, B.; Pineau, V.; Sariali, E.H.; Taouachi, R.; Isenberg, T.; Dragicevic, P. Reducing Affective Responses to Surgical Images through Color Manipulation and Stylization. Proceedings of the Joint Symposium on Computational Aesthetics, Sketch-Based Interfaces and Modeling, and Non-Photorealistic Animation and Rendering; ACM., Ed.; ACM/Eurographics, ACM: Victoria, Canada, 2018; pp. 4:1–4:13. doi:10.1145/3229147.3229158.
67. Besançon, L.; Issartel, P.; Ammi, M.; Isenberg, T. Hybrid Tactile/Tangible Interaction for 3D Data Exploration. *IEEE Transactions on Visualization and Computer Graphics* **2017**, *23*, 881–890. doi:10.1109/TVCG.2016.2599217.
68. Fröhlich, B.; Plate, J. The cubic mouse: A new device for three-dimensional input. Proc. CHI. ACM, 2000, pp. 526–531. doi:10.1145/332040.332491.
69. Gomez, S.R.; Jianu, R.; Laidlaw, D.H. A Fiducial-Based Tangible User Interface for White Matter Tractography. Advances in Visual Computing; Springer: Berlin, Heidelberg, 2010; pp. 373–381. doi:10.1007/978-3-642-17274-8_37.
70. Hinckley, K.; Pausch, R.; Goble, J.C.; Kassell, N.F. A survey of design issues in spatial input. Proc. UIST; ACM, , 1994; pp. 213–222. doi:10.1145/192426.192501.
71. Sousa, M.; Mendes, D.; Paulo, S.; Matela, N.; Jorge, J.; Lopes, D.S.o. VRRRRoom: Virtual Reality for Radiologists in the Reading Room. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; ACM: New York, NY, USA, 2017; CHI '17, pp. 4057–4062. doi:10.1145/3025453.3025566.
72. Sultanum, N.; Somanath, S.; Sharlin, E.; Sousa, M.C. "Point it, Split it, Peel it, View it": Techniques for Interactive Reservoir Visualization on Tabletops. Proc. ITS; ACM: New York, 2011; pp. 192–201. doi:10.1145/2076354.2076390.
73. Bacchetti, P. Current sample size conventions: Flaws, harms, and alternatives. *BMC Medicine* **2010**, *8*, 17. doi:10.1186/1741-7015-8-17.
74. Bravo, G.; Grimaldo, F.; López-Iñesta, E.; Mehmani, B.; Squazzoni, F. The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature communications* **2019**, *10*, 322. doi:10.1038/s41467-018-08250-2.
75. Analysing Likert scale/type data. https://www.st-andrews.ac.uk/media/capod/students/mathssupport/Likert. Accessed: 2019-06-15.

76. Stevens, S.S. On the Theory of Scales of Measurement. *Science* **1946**, *103*, 677–680, [https://science.sciencemag.org/content/103/2684/677.full.pdf]. doi:10.1126/science.103.2684.677.
77. Sauro, J. Can you take the mean of ordinal data? https://measuringu.com/mean-ordinal/#. Accessed: 2019-06-15.
78. Lewis, J.R. Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *International Journal of Human–Computer Interaction* **2002**, *14*, 463–488. doi:10.1080/10447318.2002.9669130.
79. Lewis, J.R. Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human–Computer Interaction* **1993**, *5*, 383–392. doi:10.1080/10447319309526075.
80. Sauro, J.; Lewis, J.R. *Quantifying the user experience: Practical statistics for user research, Chapter 9*; Morgan Kaufmann, 2016.
81. Lord, F.M. On the Statistical Treatment of Football Numbers. **1953**. doi:https://psycnet.apa.org/doi/10.1037/h0063675.