RESEARCH   **Open Access**

Open up: a survey on open and non-anonymized peer reviewing

Lonni Besançon[1,2]* , Niklas Rönnberg[1], Jonas Löwgren[1], Jonathan P. Tennant[3,4,5]^ and Matthew Cooper[1]

Abstract

Background: Our aim is to highlight the benefits and limitations of open and non-anonymized peer review. Our argument is based on the literature and on responses to a survey on the reviewing process of alt.chi, a more or less open review track within the so-called Computer Human Interaction (CHI) conference, the predominant conference in the field of human-computer interaction. This track currently is the only implementation of an open peer review process in the field of human-computer interaction while, with the recent increase in interest in open scientific practices, open review is now being considered and used in other fields.

Methods: We ran an online survey with 30 responses from alt.chi authors and reviewers, collecting quantitative data using multiple-choice questions and Likert scales. Qualitative data were collected using open questions.

Results: Our main quantitative result is that respondents are more positive to open and non-anonymous reviewing for alt.chi than for other parts of the CHI conference. The qualitative data specifically highlight the benefits of open and transparent academic discussions. The data and scripts are available on https://osf.io/vuw7h/, and the figures and follow-up work on http://tiny.cc/OpenReviews.

Conclusion: While the benefits are quite clear and the system is generally well-liked by alt.chi participants, they remain 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.

Keywords: Peer review, Open science

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]). With the major changes

currently happening in scholarly communication systems, there is now a strong imperative for those who manage the peer review process to be absolutely clear about their policies and, where possible, upon what evidence such policies are based [5].

The names of these different variations can be confusing for researchers. While "open review" has often been used in the past to mean "non-anonymized" reviews (e.g., [6, 7]), we will use "open review" to refer to all reviews that are publicly available, whether anonymous or signed. Classical single/double-blind reviewing is held in high regard within scientific communities and is often considered as the gold standard for assessing the validity of research communications [1–3, 8–11]. Despite the criticism it sometimes incurs [12–18], peer review is still considered to be the "best that we have" [18] and only a

*Correspondence: lonni.besancon@gmail.com
^Deceased
[1] Linköping University, Norrköping, Sweden
[2] Université Paris Sud, Orsay, France
Full list of author information is available at the end of the article

few broad-scale attempts have been made to address the numerous issues with the current system, especially in human-computer interaction.

The alt.chi conference track, however, is an exception. It is a track within the annual Computer Human Interaction (CHI) conference, which is the predominant conference in the field of human-computer interaction. 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 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 2019 reverted to the original open, volunteer-based and non-anonymous system.

In this article, our aim is 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.

Even though this paper is based on a study of a specific conference track within a specific discipline, the outcomes of the study are easily transferable to other disciplines. The questions used in the survey are not specific in any way to the discipline or the conference, only to the nature of the review process and some of the alternatives which could be used.

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 research fields.

Concerns with peer reviewing

While being almost as old as scholarship itself [19–21], peer review was only slowly formally introduced and established as the norm across the scholarly literature. In fact, one anecdote describes how Einstein chose to publish one of his papers 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 [19, 22]. While it is now well-established, peer review has often been criticized. Recent concerns include, but are not limited to (for more, see e.g., [18] or [23]), the lack of adequate training of reviewers, leading to them being unable to detect even major methodological errors [24]; the overall duration of the reviewing process which slows down progress in the scientific community [25, 26];

the unreliability of the assessments made by reviewers [27, 28]; the fact that interesting or important discussions and mitigation points highlighted by the review process are often not made accessible to other researchers [23]; that the review process is unable to prevent malicious, biased, or indifferent reviewers [14]; and that reviewers rarely receive appropriate credit for their reviews [23]. Noteworthy previous work has concluded that reviewers typically agree on a submitted manuscript at levels only slightly above chance [27] and that the current system of having two or three reviewers is unlikely to do much better than a lottery, based on mathematical modeling [29].

With respect to the CHI conference, Jansen et al. [30] conducted a survey of 46 CHI authors to determine what they value in the reviews they received in 2016. 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 to be unreasonable or insufficiently detailed. Jansen et al. 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. The authors finally argued that non-public reviews make it hard to gather data to evaluate the peer review process and added that it 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. [31] who argued that the sharing of review data could both encourage and help reward reviewers.

Types of peer review

Previous work has already investigated and attempted to summarize the main arguments for and against blinding, reciprocal or not, during peer review [6, 32, 33]. The four available and most commonly investigated options are zero-blind, single-blind, double-blind, and triple-blind. In a zero-blind system, authors, reviewers, and editors are aware of everyone's identities (although the authors usually discover the identity of their reviewers only after the reviews are made available). In a single-blind system, only the identities of the reviewers are hidden from the authors, whereas double-blind systems also hide the identities of authors from the reviewers. In a triple-blind system, even the editor is blinded to the authors' identities. It is sometimes believed that science benefits from increasing the level of anonymity.

Indeed, double-blind reviews have been shown by past research to be generally better than single-blind reviews [34–38]. It is thought to reduce reviewers' biases [35, 36, 38] and to increase the number of accepted papers with female first authors in ecology or evolution journals

[34] and seems to be preferred by both authors and reviewers [37]. Baccheli and Beller [39] showed that, despite the inherent costs of double-blind reviewing (e.g., difficulty for authors to blind papers and difficulty for reviewers to judge how incremental the work is), less than one third of the surveyed software engineering community disagreed with a switch from single-blind reviewing to double-blind reviewing. Prechelt et al. [16] investigated the perception of peer reviewing in the same community and reported that only one 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 [40]. Part of the literature argues that hiding their identity leads to better and less biased reviews [41–43], while it would seem that several large-scale studies do not support such claims [44–47]. Still, anonymizing authors appears to be one of the best solutions to address the known biases in research communities against female scientists and to increase the overall diversity of researchers engaged in the process [48–50].

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

Towards (anonymous) open peer review

With all the recent publicity surrounding open research and open access publishing, it might seem that open peer reviewing is a relatively new idea. However, journals practising open reviews have existed since at least the 1990s [51] and the possible benefits of open peer reviews have been widely discussed in the literature (e.g., [52]). The sharing of review reports in one form or another actually even goes back to the origins of peer review itself [53]. The term "open review" is, however, loosely used and encompasses several elements [18, 54] that should be distinguished [55]: open identities, open reports, open participation, open interaction, open pre-review manuscripts, open final-version commenting, and use of 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 available evidence to them about what works best.

Jones [56] argued that anonymization could be detrimental because reviewers could act without fear of sanctions and suggested that reviews should be signed. This conclusion was later supported by Shapiro [57]. There are many variations on anonymity [23]. For example, the identities of reviewers could be revealed only on published papers while reviewers of rejected papers maintain their anonymity (as is the current practice in *Frontiers in Neuroscience* [58]), 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 [59] while *BMC Public Health* (and other BMC series) has made publication of signed reviews standard practice [60]. 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* [61].

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. [62] 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 duration of the reviewing process was found to be longer [52, 63].

Method

Within human-computer interaction, 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 from the primary submission process a second chance through another round of peer reviewing with new reviewers. Over the years, it has changed many times to include an 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, including but not limited to authors and other

reviewers, can see and respond to them until the system closes.

- The system is closed. The alt.chi conference committee decides which submissions to accept, and these accepted submissions are presented at the conference. In some cases, authors are asked to attach the reviews and discussions obtained during the process to the manuscript that will be published in the conference proceedings.

To better understand the advantages and limitations of such a review process in the human-computer interaction community, we asked previous authors to complete a short https://goo.gl/forms/ZPc1y4cin32NFZc43 on the reviewing system that was in place at alt.chi. We report the survey using the CHERRIES reporting guidelines [64]. The survey is an open survey targeted at previous alt.chi authors and reviewers or chairs.

Administration
The survey, according to our institution's rules, does not need to be approved by an IRB, but participants were informed about the purpose of the survey and its approximate completion time before they started answering it. The only personal information collected were the participants' email addresses in order to inform them of the results of the study. They are stored in a separate file that only the authors can access. In addition to this, when the participants were done completing the survey, we gave them the opportunity to tell us if they did not want us to use their data and their answers (which occurred for just one participant whose answers were therefore discarded). The survey was presented as a Google form, and participation was voluntary. No incentives were directly offered, but we provided the opportunity to inform participants about the results of the study. The survey was distributed over five different pages. We did not implement a strategy to avoid multiple entries from a single individual, relying on researchers' understanding of basic survey concepts and the importance of integrity when conducting such surveys.

Recruitment
We first gathered the contact information of at least the first author of every accepted alt.chi paper from 2010 to 2018. We could not extract more information about the process (e.g., the number of submissions per year or the number of reviews they received) since the data are not available. When we believed that the first author of a publication could have already been the first author of another 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 (in total 328 emails, 20 of which received direct Mail Delivery Errors, possibly

because the authors changed their affiliations). Some of the authors we contacted have been involved in the organization of alt.chi before, and we know for sure that one of them replied to the survey (because data collection was anonymous, when respondents did not provide an email address, we cannot know whether or not they had been organizers/chairs). 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, but closed to new responses. The Google form was accepting answers between December 3 and December 17, 2018, i.e., for a total duration of 14 days.

Design and analysis
The survey comprised different categories of questions. The first category was about the person's point of view as an author (Appendix 1). The second explored the person's point of view as an alt.chi reviewer (Appendix 2). A final category (Appendix 3) 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. In the last two questions, we also sought to gather additional comments about peer review and the questionnaire itself. All questions except the final two were mandatory. The analysis was not preregistered.

Response rate and sample size
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. If we do not consider the advertisement made on social media, our survey had a response rate of 9.7%.

While such a low number of respondents could be potentially seen as problematic, it appears through the literature that, in order to gather subjective measures and opinions, it can be enough. Indeed, Isenberg et al. [65] showed that, on average, between 1 and 5 participants are used in evaluation of research projects, while Caine [66] showed that among all CHI papers published in 1 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 [67]: 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 scales, 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., [68–74]), and sometimes with numbers as low as one (e.g., [70]) or two (e.g., [71]).

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" [75] but instead the relationship between the value of a study and the size of the sample incrementally increases with each additional participant [75].

Qualitative analysis

To limit interpretational biases when analyzing the answers to open-ended questions, one of the five present authors did a first pass to categorize each comment. Two other authors used these categories to classify the 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/. Some participants gave responses which were more appropriate to the view of reviewers when asked about experiences as authors, and vice versa. Where this was apparent, the authors corrected this in the considerations of the data.

Results

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 of 30, 86.7%) 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. Most respondents (26 of 30, 86.7%) 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. The final two questions obtained a response rate of 11/30 (which is reduced to 9/30 if we consider that two participants simply stated they had no additional comment) and 9/30 (similarly 8/30 with the statement of no additional comment).

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

Concerning the alt.chi process (before CHI2018) in particular, respondents highlighted that the reviewing could simply be a popularity contest, which in the end made individual reviews less relevant (7 of 30 respondents, 23.3%). One respondent replied that the "main limitation in my mind, is that when the reviewing is public the process might become a kind of popularity contest, or a test of

who can bring the most supporters to the table." Furthermore, in the alt.chi process, papers deemed uninteresting had less chance of acceptance as they would receive less reviews (4 of 30 respondents, 13.3%), and the limits of the invite-to-review (i.e., open participation) system were pointed out, as authors could invite friends to review (2 respondents, 6.7%).

Overall, respondents praised the discussions that the open review process of alt.chi (before CHI2018) brought, which is an advantage for both authors (13 of 30 respondents, 43.3%) and reviewers (14 of 30 respondents, 46.7%) and can also stimulate the discussions between reviewers (3 of 30 respondents, 10%). For example, one respondent stated that the open review process has the "[p]otential for discussion and critique between authors and reviewers during the review process, rather than the summative evaluation (accept / reject) in the full papers track." The added transparency in the reviewing process was praised (5 of 30 respondents, 16.7%) as a benefit for authors as it helps them understand the comments from reviewers (2 of 30 respondents, 6.7%) and can reduce the *cite-me* effect (1 of 30 respondent, 3.3%). One respondent replied that "transparency is always welcome. I think reviewers are more constructive if their reviews are non-anonymous. Also the potential risk of reviewers asking 'please quote me' disappears." The respondents mentioned that reviewers used a more polite tone (4 of 30 respondents, 13.3%), that the open review process fosters future collaborations as authors can directly contact reviewers and vice versa (2 of 30 respondents, 6.7%), and that the more diverse set of reviewers could also lead to interesting discussions (2 of 30 respondents, 6.7%). The respondents also 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 of 30 respondents, 20%) and reviewers (8 of 30 respondents, 26.7%). As one respondent stated: "An actual discussion was possible [i.e. before CHI2018], and people mostly commented only if they actually had a well-founded opinion." Interestingly, three respondents mentioned that signing reviews was a good way to receive credit for their work.

Considering open/public and non-anonymized reviewing, some respondents expressed concerns that reviewers might fear being truly critical and, consequently, self-censor their reviews (14 of 30 respondents, 46.7%) and that an author's reputation could possibly directly influence the reviewer and the decision on the submission (4 of 30 respondents, 13.3%, as a limitation for authors, 2 of 30 respondents for reviewers, 6.7%). One respondent stated: "I think there is a lot of self-censorship and trying not to step on more senior people's toes." Finally, negative reviews, even if well-founded, could generate animosity

and result in retaliation with respect to future submissions by the reviewer (4 of 30 respondents, 13.3%).

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 b. 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 [76]) in addition to means and medians. While the use of means for ordinal values has been initially slightly advocated against [77] and is still highly controversial [78], it appears in the literature that it is nonetheless highly used [79], useful to present [77, 78, 80, 81], and potentially even more useful than medians [80, 82].

The results in Fig. 1a and b highlight the openness and interest towards an open and non-anonymous review process that was already suggested by our qualitative results. Indeed, 23 respondents (of 30) 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 the whole CHI conference, 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 respondents would not consider it: 18 gave scores of 1 or 2 (mean = 2.23, median = 2).

Discussions

Our study's qualitative and quantitative results suggest that the respondents have a general interest towards open and non-anonymous review processes. However, more than half of the respondents would nevertheless not consider signed reviews for other tracks of CHI. This might be due to the risk of retaliation for reviewers of a rejected paper, as mentioned by some of the respondents and echoing findings from previous work (e.g., [10, 55, 83]). 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. Still, such alternatives are rarely used, and we hypothesize that they were probably not considered by most of our respondents (though future work could investigate this aspect further). Nonetheless, the reluctance to sign reviews for other CHI tracks contrasts with the rapidly growing number of journals that are using non-anonymous and public reviews (see, e.g., some of the BMC series [60] and the https://transpose-publishing.github.io site for a complete list).

The respondents indicated the limitations of the invite-to-review system, such as asking friends to review or turning the process into a popularity contest. Such problems are, however, not inherent to open and non-anonymous reviewing but rather emerge from the specific alt.chi
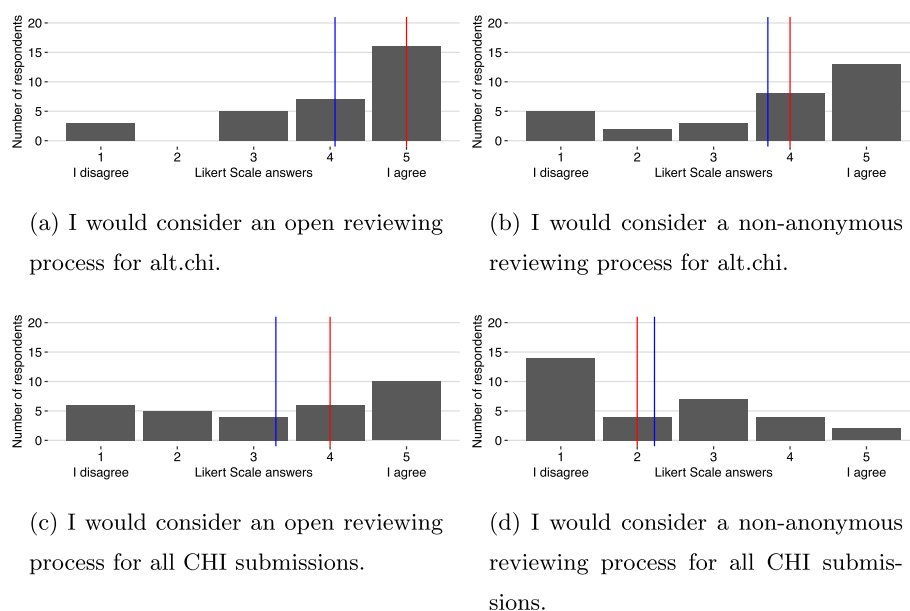

(a) I would consider an open reviewing process for alt.chi.

(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.

Fig. 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

implementation. An obvious improvement would to have a fixed number of assigned reviewers while still keeping the system open and non-anonymous.

The notion that reviewers might use a more polite tone when doing open reviews mirrors previous literature findings [52, 63], and it seems reasonable to assume that a more polite tone also could foster future collaborations between researchers. Some respondents pointed out that open reviews would make reviewer comments more justified as the reviewers would be directly accountable for their reviews (see also Jansen et al.'s [30] findings).

Limitations and future work

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. In addition, it should be noted that, even with our biased sample of previous alt.chi authors and reviewers, our results indicate that many of them consider that reviewers should remain anonymous in other CHI tracks or SIGCHI venues. This therefore suggests that the level of acceptance for broadening this practice even among researchers who have participated in open peer review before is quite low. We believe that this is a particularly interesting challenge that the open science community has to take into account: exposure to, and acceptance of, open systems or practices in specific contexts does not necessarily translate into other contexts.

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. This could then be compared to peer review at conferences for other constituencies within the wider software engineering community.

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. Our 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 and non-anonymous process. For instance, having a fixed number of assigned reviewers could solve many of the issues identified in 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. It would also seem that the alt.chi community fears that the implementation of non-anonymous reviews in more prestigious venues could lead to issues such as 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 human-computer interaction community is using it only at alt.chi where accepted papers count only as extended abstracts rather than full archival publications in the proceedings of the conference. Indeed, our empirical findings seem to support the old adage that "double-blind peer review is the worst academic QA system, except for all the others." It is nevertheless our hope that our work can contribute to further discussions on open peer reviewing processes and to experimentation with such processes in other academic venues. The small-scale survey implemented here could easily be adapted to help other scientific communities further understand and optimize their own peer review processes.

Appendix 1
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 2
Questions as a reviewer

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

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? (Open)

Appendix 3
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)

18 Do you have any additional comments on the questionnaire itself? (Open)

Abbreviations
CHI: The main venue for work done in human-computer interaction

Authors' contributions
Lonni Besançon: conceptualization, data curation, investigation, project administration, validation, software, visualization, and writing of the original draft. Niklas Rönnberg: writing of the original draft and data curation. Jonas Löwgren: writing of the original draft, writing, review, and editing. Jonathan P. Tennant: writing, review, and editing. Matthew Cooper: writing of the original draft and data curation. The author(s) read and approved the final manuscript.

Availability of data and materials
The data and scripts are available on https://osf.io/vuw7h/, and the figures and follow-up work on http://tiny.cc/OpenReviews.

Ethics approval and consent to participate
As per the first author's institution, online surveys requiring no sensitive or identifiable information are not subject to ethics approval.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Linköping University, Norrköping, Sweden. [2]Université Paris Sud, Orsay, France. [3]Southern Denmark University Library, Campusvej 55, 5230 Odense, Denmark. [4]Center for Research and Interdisciplinarity, Universite de Paris, Rue Charles V, Paris, France. [5]Institute for Globally Distributed Open Research and Education, Bali, Indonesia.

References
1.   Morgan PP. Anonymity in medical journals. Can Med Assoc J. 1984;131(9): 1007–8.
2.   Pierson CA. Peer review and journal quality. J Am Assoc Nurse Pract. 2018;30(1):.
3.   Wilson JD. Peer review and publication. Presidential address before the 70th annual meeting of the American Society for Clinical Investigation, San Francisco, California, 30 April 1978. J Clin Investig. 1978;61(6): 1697–701. https://doi.org/10.1172/JCI109091.
4.   Largent EA, Snodgrass RT. In: Robertson CT, Kesselheim AS, editors. Chapter 5 Blind peer review by academic journals: Academic Press; 2016, pp. 75–95. https://doi.org/10.1016/b978-0-12-802460-7.00005-x.
5.   Klebel T, Reichmann S, Polka J, McDowell G, Penfold N, Hindle S, Ross-Hellauer T. Peer review and preprint policies are unclear at most major journals. BioRxiv. 2020. https://doi.org/10.1101/2020.01.24.918995.
6.   Pontille D, Torny D. The blind shall see! The question of anonymity in journal peer review. Ada: J Gender New Media Technol. 2014;4:. https://doi.org/10.7264/N3542KV.
7.   Ross-Hellauer T. What is open peer review? A systematic review. F1000Research. 2017;6:. https://doi.org/10.12688/f1000research.11369.2.
8.   Baggs JG, Broome ME, Dougherty MC, Freda MC, Kearney MH. Blinding in peer review: the preferences of reviewers for nursing journals. J Adv Nurs. 2008;64(2):131–8. https://doi.org/10.1111/j.1365-2648.2008.04816.x.
9.   Haider J, Åström F. Dimensions of trust in scholarly communication: problematizing peer review in the aftermath of John Bohannon's "Sting" in science. J Assoc Inf Sci Technol. 2016;68(2):450–67. https://doi.org/10.1002/asi.23669.
10. Mulligan A, Hall L, Raphael E. Peer review in a changing world: an international study measuring the attitudes of researchers. J Am Soc Inf Sci Technol. 64(1):132–61. https://doi.org/10.1002/asi.22798.
11. Moore S, Neylon C, Eve MP, O'Donnell DP, Pattinson D. "Excellence R Us": university research and the fetishisation of excellence. Palgrave Commun. 2017;3:16105. https://doi.org/10.1057/palcomms.2016.105.
12. Armstrong JS. Peer review for journals: evidence on quality control, fairness, and innovation. Sci Eng Ethics. 1997;3(1):63–84. https://doi.org/10.1007/s11948-997-0017-3.
13. Baxt WG, Waeckerle JF, Berlin JA, Callaham ML. Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. Ann Emerg Med. 1998;32(3):310–7. https://doi.org/10.1016/S0196-0644(98)70006-X.
14. D'Andrea R, O'Dwyer JP. Can editors save peer review from peer reviewers?. PLOS ONE. 2017;12(10):1–14. https://doi.org/10.1371/journal.pone.0186111.

15. Hettyey A, Griggio M, Mann M, Raveh S, Schaedelin FC, Thonhauser KE, Thoß M, van Dongen WFD, White J, Zala SM, Penn DJ. Peerage of science: will it work?. Trends Ecol Evol. 2012;27(4):189–90. https://doi.org/10.1016/j.tree.2012.01.005.

16. Prechelt L, Graziotin D, Fernández DM. A community's perspective on the status and future of peer review in software engineering. Inf Softw Technol. 2018;95:75–85. https://doi.org/10.1016/j.infsof.2017.10.019.

17. 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(1151):. https://doi.org/10.12688/f1000research.12037.3.

18. Tennant JP. The state of the art in peer review. FEMS Microbiol Lett. 2018;365(19):204. https://doi.org/10.1093/femsle/fny204.

19. Baldwin M. In referees we trust?. Phys Today. 2017;70(2):44–9. https://doi.org/10.1063/pt.3.3463.

20. Baldwin M. What it was like to be peer reviewed in the 1860s. Phys Today. 2017. https://doi.org/10.1063/PT.3.3463.

21. Spier R. The history of the peer-review process. Trends Biotechnol. 2002;20(8):357–8. https://doi.org/10.1016/S0167-7799(02)01985-6.

22. Kennefick D. Einstein versus the physical review. Phys Today. 2005;58(9):43. https://doi.org/10.1063/1.2117822.

23. Walker R, Rocha da Silva P. Emerging trends in peer review—a survey. Front Neurosci. 2015;9:169. https://doi.org/10.3389/fnins.2015.00169.

24. 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(7441):673. https://doi.org/10.1136/bmj.38023.700775.AE.

25. Bornmann L, Daniel H-D. How long is the peer review process for journal manuscripts? A case study on Angewandte Chemie International Edition. CHIMIA Int J Chem. 2010;64(1):72–7. https://doi.org/10.2533/chimia.2010.72.

26. Benos DJ, Bashari E, Chaves JM, 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 RG, Zotov A. The ups and downs of peer review. Adv Physiol Educ. 2007;31(2):145–52. https://doi.org/10.1152/advan.00104.2006.

27. Kravitz RL, Franks P, Feldman MD, Gerrity M, Byrne C, Tierney WM. Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care?. PLOS ONE. 2010;5(4):1–5. https://doi.org/10.1371/journal.pone.0010072.

28. Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. Cogn Therapy Res. 1977;1(2):161–75. https://doi.org/10.1007/BF01173636.

29. Herron DM. Is expert peer review obsolete? A model suggests that post-publication reader review may exceed the accuracy of traditional peer review. Surg Endosc. 2012;26(8):2275–80. https://doi.org/10.1007/s00464-012-2171-1.

30. Jansen Y, Hornbaek K, Dragicevic P. What did authors value in the CHI'16 reviews they received?. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; 2016. https://doi.org/10.1145/2851581.2892576.

31. Squazzoni F, Grimaldo F, Marušić A. Publishing: journals could share peer-review data. Nature. 2017;546:352.

32. Jubb M. Peer review: the current landscape and future trends. Learn Publ. 2016;29(1):13–21. https://doi.org/10.1002/leap.1008.

33. Snodgrass R. Single- versus double-blind reviewing: an analysis of the literature. SIGMOD Rec. 2006;35(3):8–21. https://doi.org/10.1145/1168092.1168094.

34. Budden AE, Tregenza T, Aarssen LW, Koricheva J, Leimu R, Lortie CJ. Double-blind review favours increased representation of female authors. Trends Ecol Evol. 2008;23(1):4–6. https://doi.org/10.1016/j.tree.2007.07.008.

35. Jefferson T, Godlee F. Peer Review in Health Sciences. London: BMJ Books; 2003.

36. Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. JAMA. 1994;272(2):96–7. https://doi.org/10.1001/jama.1994.03520020022005.

37. Regehr G, Bordage G. To blind or not to blind? What authors and reviewers prefer. Med Educ. 2006;40(9):832–9. https://doi.org/10.1111/j.1365-2929.2006.02539.x.

38. Ross JS, Gross CP, Desai MM, Hong Y, Grant AO, Daniels SR, Hachinski VC, Gibbons RJ, Gardner TJ, Krumholz HM. Effect of blinded peer review on abstract acceptance. JAMA. 2006;295(14):1675–80. https://doi.org/10.1001/jama.295.14.1675.

39. Bacchelli A, Beller M. Double-blind review in software engineering venues: the community's perspective. In: 2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C); 2017. p. 385–96. https://doi.org/10.1109/ICSE-C.2017.49.

40. Tennant JP. The dark side of peer review. In: EON; 2017. p. 2–4. https://doi.org/10.18243/eon/2017.10.8.1.

41. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review: a randomized trial. JAMA. 1990;263(10):1371–6. https://doi.org/10.1001/jama.1990.03440100079012.

42. Baggs JG, Broome ME, Dougherty MC, Freda MC, Kearney MH. Blinding in peer review: the preferences of reviewers for nursing journals. J Adv Nurs. 2008;64(2):131–8.

43. Weicher M. Peer review and secrecy in the "information age". Proc Am Soc Inf Sci Technol. 2008;45(1):1–12.

44. Isenberg SJ, Sanchez E, Zafran KC. The effect of masking manuscripts for the peer-review process of an ophthalmic journal. Br J Ophthalmol. 2009;93(7):881–4. https://doi.org/10.1136/bjo.2008.151886.

45. Justice AC, Cho MK, Winker MA, Berlin JA, Rennie D, the PEER Investigators. Does masking author identity improve peer review quality? A randomized controlled trial. JAMA. 1998;280(3):240–2. https://doi.org/10.1001/jama.280.3.240.

46. Lee CJ, Sugimoto CR, Zhang G, Cronin B. Bias in peer review. J Am Soc Inf Sci Technol. 2013;64(1):2–17. https://doi.org/10.1002/asi.22784.

47. 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(3):234–7. https://doi.org/10.1001/jama.280.3.234.

48. Darling ES. Use of double-blind peer review to increase author diversity. Conserv Biol. 2015;29(1):297–9. https://doi.org/10.1111/cobi.12333.

49. Helmer M, Schottdorf M, Neef A, Battaglia D. Research: Gender bias in scholarly peer review. eLife. 2017;6:21718. https://doi.org/10.7554/eLife.21718.

50. Roberts SG, Verhoef T. Double-blind reviewing at evolang 11 reveals gender bias†. J Lang Evol. 2016;1(2):163–7. https://doi.org/10.1093/jole/lzw009.

51. Parks S GS. Tracking global trends in open peer review. https://www.rand.org/blog/2017/10/tracking-global-trends-in-open-peer-review.html. Accessed 15 June 2020.

52. Walsh E, Rooney M, Appleby L, Wilkinson G. Open peer review: a randomised controlled trial. Br J Psychiatry. 2000;176(1):47–51. https://doi.org/10.1192/bjp.176.1.47.

53. Csiszar A. Peer review: troubled from the start. Nat News. 2016;532(7599):306. https://doi.org/10.1038/532306a.

54. Ross-Hellauer T, Schmidt B, Kramer B. Are funder open access platforms a good idea?. PeerJ Preprints. 2018;6:26954–1. https://doi.org/10.7287/peerj.preprints.26954v1.

55. Ross-Hellauer T, Deppe A, Schmidt B. Survey on open peer review: attitudes and experience amongst editors, authors and reviewers. PLOS ONE. 2017;12(12):0189311.

56. Jones R. Rights, wrongs and referees. N Sci. 1974;61(890):758–9.

57. Shapiro BJ. A culture of fact: England, 1550-1720. Ithaca: Cornell University Press; 2003.

58. in Neuroscience F. Frontiers in Neuroscience Review System. https://www.frontiersin.org/about/review-system.Accessed 15 June 2020.

59. PeerJ. Policies and procedures. https://peerj.com/about/policies-and-procedures/. Accessed 15 June 2020.

60. Health BP. Peer review policy. https://bmcpublichealth.biomedcentral.com/submission-guidelines/peer-review-policy.

61. of Bioethics TAJ. Standards for manuscript submission general information. http://www.bioethics.net/wp-content/uploads/2012/02/Standards-for-Manuscript-Submission.pdf?x63245. Accessed 15 June 2020.

62. Pucker B, Schilbert H, Schumacher SF. Integrating molecular biology and bioinformatics education. Preprints 2018. 2018. https://doi.org/10.20944/preprints201811.0183.v1. Accessed 15 June 2020.

63. Snell L, Spencer J. Reviewers' perceptions of the peer review process for a medical education journal. Med Educ. 2005;39(1):90–7. https://doi.org/10.1111/j.1365-2929.2004.02026.x.

64. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-surveys (CHERRIES). J Med Internet Res. 2004;6(3):34.

65. Isenberg T, Isenberg P, Chen J, Sedlmair M, Möller T. A systematic review on the practice of evaluating visualization. IEEE Trans Vis Comput Graph. 2013;19(12):2818–27. https://doi.org/10.1109/TVCG.2013.126.

66. Caine K. Local standards for sample size at chi. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16. New York: ACM; 2016. p. 981–992. https://doi.org/10.1145/2858036.2858498.

67. Koeman L. How many participants do researchers recruit? A look at 678 UX/HCI studies. 2018. https://lisakoeman.nl/blog/how-many-participants-do-researchers-recruit-a-look-at-678-ux-hci-studies. Accessed 6 Jan 2019.

68. Besançon L, Semmo A, Biau DJ, Frachet B, Pineau V, Sariali EH, Taouachi R, Isenberg T, Dragicevic P. Reducing affective responses to surgical images through color manipulation and stylization. In: ACM, editor. Proceedings of the Joint Symposium on Computational Aesthetics, Sketch-Based Interfaces and Modeling, and Non-Photorealistic Animation and Rendering. Victoria: ACM; 2018. p. 4–1413. https://doi.org/10.1145/3229147.3229158. ACM/Eurographics. https://hal.inria.fr/hal-01795744.

69. Besançon L, Issartel P, Ammi M, Isenberg T. Hybrid tactile/tangible interaction for 3D data exploration. IEEE Trans Vis Comput Graph. 2017;23(1):881–90. https://doi.org/10.1109/TVCG.2016.2599217.

70. Fröhlich B, Plate J. The cubic mouse: a new device for three-dimensional input. In: Proc. CHI; 2000. p. 526–31. https://doi.org/10.1145/332040.332491. ACM.

71. Gomez SR, Jianu R, Laidlaw DH. A fiducial-based tangible user interface for white matter tractography. In: Advances in visual computing. Berlin, Heidelberg: Springer; 2010. p. 373–81. https://doi.org/10.1007/978-3-642-17274-8_37.

72. Hinckley K, Pausch R, Goble JC, Kassell NF. A survey of design issues in spatial input. In: Proc. UIST. New York; 1994. p. 213–22. https://doi.org/10.1145/192426.192501. ACM.

73. Sousa M, Mendes D, Paulo S, Matela N, Jorge J, Lopes DSo. Vrrrroom: Virtual reality for radiologists in the reading room. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17. New York: ACM; 2017. p. 4057–62. https://doi.org/10.1145/3025453.3025566.

74. Sultanum N, Somanath S, Sharlin E, Sousa MC. "Point it, split it, peel it, view it": techniques for interactive reservoir visualization on tabletops. In: Proc. ITS. New York: ACM; 2011. p. 192–201. https://doi.org/10.1145/2076354.2076390.

75. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Medicine. 2010;8(1):17. https://doi.org/10.1186/1741-7015-8-17.

76. Analysing Likert scale/type data. https://www.st-andrews.ac.uk/media/capod/students/mathssupport/Likert. Accessed 15 June 2019.

77. Stevens SS. On the theory of scales of measurement. Science. 1946;103(2684):677–80. https://doi.org/10.1126/science.103.2684.677.

78. Sauro J. Can you take the mean of ordinal data?. https://measuringu.com/mean-ordinal/#. Accessed 06 June 2019.

79. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Hum Comput Interact. 2002;14(3-4):463–88. https://doi.org/10.1080/10447318.2002.9669130.

80. Lewis JR. Multipoint scales: mean and median differences and observed significance levels. Int J Hum Comput Interact. 1993;5(4):383–92. https://doi.org/10.1080/10447319309526075.

81. Sauro J, Lewis JR. Quantifying the user experience: practical statistics for user research, Chapter 9. Burlington: Morgan Kaufmann; 2016.

82. Lord FM. On the statistical treatment of football numbers. 1953. https://doi.org/10.1037/h0063675.

83. 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. Nat Commun. 2019;10(1):322. https://doi.org/10.1038/s41467-018-08250-2.

Publisher's Note