

What Did Authors Value in the CHI'16 Reviews They Received?

Yvonne Jansen, Kasper Hornbaek, Pierre Dragicevic

► **To cite this version:**

Yvonne Jansen, Kasper Hornbaek, Pierre Dragicevic. 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, May 2016, San Jose, United States. pp.596 - 608, 2016, <10.1145/2851581.2892576>. <hal-01405042>

HAL Id: hal-01405042

<https://hal.inria.fr/hal-01405042>

Submitted on 29 Nov 2016

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

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

What Did Authors Value in the CHI'16 Reviews They Received?

Yvonne Jansen

University of Copenhagen
yvja@di.ku.dk

Pierre Dragicevic

Inria, France
pierre.dragicevic@inria.fr

Kasper Hornbæk

University of Copenhagen
kash@di.ku.dk

Abstract

Peer-review is key to assessing work in HCI conferences. The content and process of peer-review, and how it moves scholarship forward or impedes it, are much discussed but little data is available. We provide initial data from surveying 46 authors who submitted papers and notes to CHI 2016, and asking them what they found helpful and unhelpful in their reviews. Responses were overall positive, and showed that authors appreciated encouragement, ideas for related work, and seeing their work fairly assessed. At the same time, some authors commented that reviews may not be inclusive of new approaches, may contain insufficient details, and occasionally seem unreasonable. They also noted issues specific to the rebuttal process. We discuss how instructions for reviewers could be improved, and link our findings to ongoing debates on peer review.

Author Keywords

Peer review; user studies; CHI; surveys

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]

CHI'16 Extended Abstracts, May 07 - 12, 2016, San Jose, CA, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4082-3/16/05... \$15.00
Author version. For the definitive version see <http://dx.doi.org/10.1145/2851581.2892576>

Introduction

What do people who submit to CHI value to see in a peer review of their submission? What do they like and dislike? These are common topics of discussion, both in anticipation of the reviews and after receiving them. Whereas personal anecdotes and opinions are legion, little factual information is available. Opinions on what is appropriate in a review likely differ across people, particularly in a multicultural community such as CHI. Also, reviews and unfiltered opinions are rarely shared outside local peer groups, thus it is difficult to get a sense of what authors really value in the reviews they receive, and what constitutes an effective review. Overall, it is hard for authors to know what to expect from a review, difficult – in particular for junior reviewers – to know what to put into a review, and unclear for CHI as a whole how to create effective reviewing guidelines.

In this article we summarize data collected from 46 authors who shared their experiences (sometimes with quotes from reviews) through an “unofficial CHI reviewing survey”. In contrast to the official CHI reviewing survey, the aim of our survey is to start an *open* discussion of reviews with the community, based on data and facts. We hope our survey will inspire the collection of more data to help reviewers craft better reviews and conference chairs craft more effective reviewing guidelines.

Method

We conducted an online survey on the content and perceived quality of CHI '16 reviews, including a part focusing on reporting strategies for user studies. We asked questions such as “What is your overall impression of the review quality for your submission?”, “In your [study] analyses, what quantities did you report in numerical form?”, or “Which aspects of your study received positive comments from reviewers?”. The full survey with all questions is available at tinyurl.com/unofsurv.

Three questions invited respondents to share review quotes and comment on them. A link was provided to a text where we discuss the legal, ethical and practical implications of sharing received reviews, with links to external sources (mostly discussion forums). The full text is available at tinyurl.com/sharereviews. We stated that the dataset will be shared publicly, and gave respondents the option to keep their review quotes and/or personal comments private.

We advertised the survey through snowballing (online social networks and emails) starting half an hour after CHI submitters had received their final reviews (16 Dec 2015). We tried to reach out to as many communities as possible by emailing researchers from different cities in Americas, Europe, and Asia. The survey remained open for a period of 20 days, during which all respondents (and their co-authors) could edit or retract their responses.

As survey responses were not randomly sampled from all CHI submissions (due to self-selection and possibly community bias), we do not report inferential statistics but summary statistics and a qualitative analysis of responses.

Quantitative Overview of Responses

We received 52 responses submitted by 46 respondents. Each response concerns a single CHI submission, and authors of multiple submissions could respond multiple times.

Submission Types

The responses included 36 full papers (69%), 14 notes, and 2 unspecified. The most common CHI contribution types (following the typology listed on the [CHI website](#)) were *Understanding Users* (44%) and *Development or Refinement of Interface Artifacts or Techniques* (21%). Of the 52 responses, 20 (38%) were accepted (2 with shepherding), 31 (60%) were rejected, and 1 was unspecified. Responses were thus slightly biased towards accepted submissions¹.

Perceived Review Quality

Authors' overall perception of review quality was rather positive but very diverse, ranging from exceptionally good to exceptionally poor (see Figure 1). Reviews of rejected submissions tended to be judged more negatively than reviews of accepted submissions. Nevertheless, reviews of rejected submissions were occasionally well-received, and some authors of accepted papers felt they got poor reviews.

User Studies

Forty-four out of 52 responses (85%) reported a user study. Of these, 8 were quantitative, 9 were qualitative, 26 used mixed methods, and 1 was unspecified. Studies involved from 5 to 1618 participants in total (median 30), summed

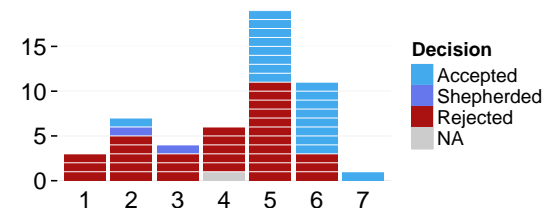


Figure 1: Perceived review quality for each response ($n = 52$), from exceptionally poor (1) to exceptionally good (7).

¹The acceptance rate for papers at CHI'16 was 23.4%.

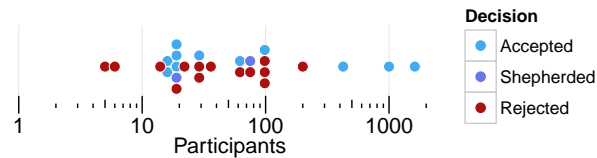


Figure 2: Total number of participants per study ($n = 44$).

across all experiments pilots included (see Figure 2). Sample sizes were thus extremely diverse and did not predict acceptance, except perhaps in the extremes.

Among the 34 studies where at least some quantitative data were reported (this includes mixed methods), diverse statistical analysis and reporting strategies were used (Figure 4). It appears that papers were accepted and rejected across all categories. Bayesian analysis options were offered in the form but no respondent selected them. The qualitative methods used were diverse but a further analysis of these methods would go beyond the scope of this article.

The degree to which reviewers approved of statistical methods varied a lot across responses, according to respondents (Figure 3, x-axis). Reviewers' approval was strongly linked to paper acceptance. Authors also varied as to how much they agreed with reviewers' methodological concerns (Figure 3, y-axis). Although agreement was frequent, strong reviewer disapproval never led to author agreement.

Textual Content

Responses included a total of 74 review quotes and 46 personal comments on these quotes. Respondents agreed to make 50 (68%) of the review quotes public, and 30 (65%) of the comments public. Some comments are quoted below, and more data is available at tinyurl.com/chi16reviewdata.

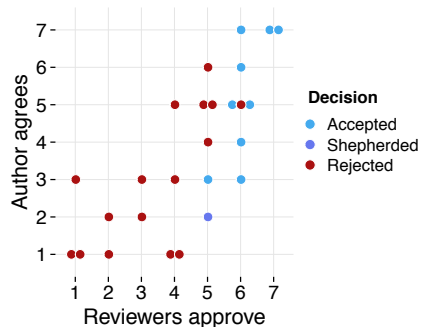


Figure 3: Scatterplot showing to what extent reviewers approve of the author's choice of data analysis and reporting strategy, and to what extent the author agrees with their criticisms ($n = 34$).

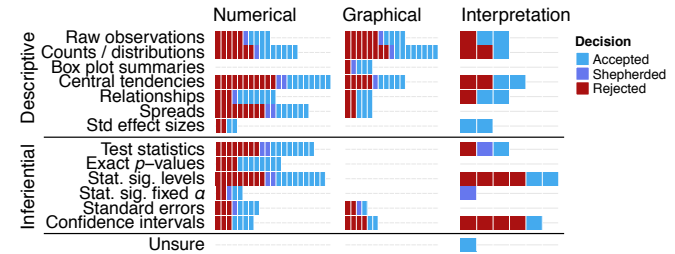


Figure 4: Quantities reported numerically and graphically (multiple responses were allowed), and quantities primarily used for interpreting results and drawing conclusions ($n = 34$).

Qualitative Analysis

Textual content was classified through affinity diagramming: we printed a summary report for each response, which we cut up in pieces and sorted thematically. We used these themes to write a code book containing 37 codes grouped in 9 categories. One author coded the entire set of quotes, comments, and suggestions, while the other two authors coded one third each. We used the frequency counts of codes as a measure for including themes in this report and provide a subjective selection of salient responses here. The following sections contain paraphrased reviewer quotes² and quotes from survey respondents which are typeset differently. Keep in mind that reviewer comments are necessarily taken out of context and may not reflect “the truth” but are only intended to give some insight into the author's perception of their reviews.

As far as possible, we tried to contrast high-quality reviewer comments with those authors considered to be of low-quality. However, review quotes which authors considered

²Due to a controversy during the alt.chi review process over whether or not publishing review quotes without reviewer permission is acceptable, we were asked by the alt.chi chairs to redact direct reviewer quotes.

high-quality are not necessarily positive in content, but might have been selected by respondents because they were particularly helpful, informative, or constructive.

General Comments on Reviews

A number of general comments, both positive and negative, were offered concerning content, form and tone of reviews.

Encouraging Tone

In the “high-quality reviews” section of the survey, ten responses included review quotes with encouraging comments. Encouraging comments could be simple and rather unspecific such as that all reviewers recognized that the paper was doing a great job. Authors particularly appreciated when reviewers not only critiqued their submission, but also expressed which parts were good or whether the authors were on to something interesting and relevant.

Detailed versus Vague Comments

Authors appreciated detailed reviews that provided clear and actionable remarks. One response highlights an exceptionally constructive review with the following explanation:

“This reviewer presented overall review / summary followed by the list of things that needs to be addressed with SUBSTANTIVE marks when needed. I really appreciated this review because it was done by a domain expert (4 on self-rating, but I think this reviewer deserves 5) and VERY thorough. Also, with SUBSTANTIVE mark, I could see what are the weakest points of my paper and figure out how to make it stronger (it also helped me set my rebuttal strategy)”

Conversely, four responses included complaints about remarks being too vague, such as the necessity to do “some” extra experiments, on which the author commented:

“This is vague, unhelpful, and could be applied to any paper. It is unclear what extra experiments this reviewer expects to see, and no suggestions were offered.”

Similarly, one response included a quote asking for a “light-weight informal study” without any suggestions how such a study might look like. The author stated that they discussed this question in the paper and the rebuttal. Thus more concrete suggestions what the reviewer would consider sufficient would have been more constructive.

One author sent us the following comment:

“... this reviewer was “sure others already did this” - but this reviewer couldn't point to a similar experiment, study or finding in literature to support his claim.”

The overall sentiment can be summarized by one respondent who suggested:

“If you say the findings are not new, please give at least one example saying where it has been done.”

Balance

Balanced reviews that acknowledge the positive while providing clear recommendations for improving the weak parts were particularly well received. One respondent appreciated the “detailed responses from referees [...] including what they found good, what they found lacking, and the recommendations offered”, while another one found a reviewer’s comment

“particularly good because it made a positive comment (having made explicit ethical issues), criticized (maybe still dangerous for patients), pointed to people with whom this should be discussed (IRB), and provided literature pointers to read about this subject.”

Another author commented:

“I think the previous comment was particularly useful because the reviewer was not only criticizing, but putting her or himself in my place and explicitly telling me how they would change it.”

Attention Paid to the Paper

Two responses lauded reviewers for having read the paper carefully, e.g., one author appreciated that “reviewers read and understood the paper”.

At the same time, eight responses included remarks about reviewers not reading the submission carefully enough. One author suggest as a guideline *“that reviewers read the paper carefully, or not review it at all?”* Another author offered the following comment:

“RTFM - where M stands for Manuscript! Some papers are trying to address a complex topic and it cannot be made simpler without dumbing down the quality of the submission - to provide scientific rigour and good scholarship the author needs to use complex arguments. If you do not have time to read the paper then do not accept to review the paper - it is worse to volunteer and do a bad job than to not volunteer at all!”

Level of Understanding

Related to the previous issue is the reviewer’s level of understanding of the submission content. While one response expressed appreciation for a reviewer’s level of understanding, seven responses complained about a lack of understanding. Here are comments from four different responses:

“Reviewer completely misunderstood the analysis. There was only one independent variable.”

“The questions R2 raised were already answered in the papers. It was clear that he/she did not understand parts of our design.”

“one reviewer saying that he can’t believe the results but obviously he did not understand the methods and another reviewer clearly not understanding the value of qualitative results”

“Twice (reviews + rebuttal) my AC did not clarify on a misunderstanding from one reviewer. The reviewer did totally misunderstand the concept while all others did not.”

Unreasonable Requests

Seven responses referred to unreasonable requests such as effectively asking authors to write a different paper. One author felt that a reviewer’s remark *“would apply to the next iteration, not this one”* while another author commented: *“Asking to expand discussion beyond the scope of our work or what our data can cover seemed too much to ask.”*

Similarly, another author commented:

“This felt like a classic “the reviewer wished we wrote a different paper than we did” review. They were well-intentioned and were clearly trying to make our work stronger, but this review was not helpful in evaluating our current submission.”

Structure and Length

Reviewers’ styles concerning structure and length of their reviews varied considerably. We highlight here a few examples which authors deemed particularly valuable.

For a rejected submission the author indicated as high quality aspects of the received reviews *“very long and detailed reviewers”*, and that *“reviewers sorted their comments to give major and minor comments”*. Another one mentioned:

“Two of the externals broke down their review into paragraphs like: Significance, Originality, Validity, Method, Presentation clarity, Previous work. I like this format, since it more encourages to cover all the goods and the bads.”

While authors generally praised long reviews, three commented negatively on reviews that were too short: *“Writing a paragraph is not considered a review at CHI”* and:

“There should be a minimum length for the reviews - it’s ridiculous to see some reviews that are one paragraph vs. 3 pages long.”

“[...] students work very hard, and they deserve more than one-liner reviews from the most respected conference in HCI.”

No author complained about a review that was too long.

Remarks Inappropriate in a Review

Four responses referred to remarks that could be deemed inappropriate in a review. For example, one reviewer said that what an author did in their submission would make the entire community look bad and should therefore not be supported. One primary AC offered a personal attack as a post-rebuttal by commenting that for a paper written by well-known authors, the number of issues was surprising.

Studies

User study methodology traditionally generates lots of disagreement between researchers, and this is why one part of our survey focused on user studies.

Study mandatory?

Only a few responses referred to the long-standing debate on whether every CHI paper needs a study (we counted one positive reference, one negative and two neutral). One author shared review excerpts where a reviewer asks for a field study to validate findings resulting from a literature review. Another reviewer wrote as a post-rebuttal comment that while Infovis published cool but unusable ideas, there should be at least for CHI some form of validation.

The same author was told from a different reviewer that the efficiency of their proposed techniques would be difficult to validate with a user study, as no status quo existed to compare against.

Reviewers occasionally asked for an extra study. One author stated that

“referees wanted far more precision than was appropriate for a study informing initial design (where we were trying to scope out the design space), and then asked for yet another study of the system we developed from that study.”

Sample Size

The number of study participants is another common point of contention. Indeed, authors and reviewers disagreed (often inconsistently across responses) on what is an “appropriate” sample size, particularly for qualitative studies.

An author offers as a suggestion for reviewers:

“Pay more attention to the contribution of the paper than criticizing small points that makes less difference in the final results such as number of participants in a qualitative study, the need of evaluation in every study.”

Statistical Reporting Methods

Ten responses referred to the use of statistical analysis and reporting methods. We received responses from authors using different approaches, where each was sometimes positively received, and sometimes negatively received.

For example, in one response, an author pointed out that reviewers did not agree with the use of estimation, i.e., reporting interval estimates without p -values:

“One of the reviewer was concerned with the methodology we used for analyzing our data. We calculated confidence intervals (bootstrapping), and presented our results using graphs. [...] We were reproached of not using inferential statistics. [...] estimation methods ARE inferential statistics, so the argument per se is unfounded and fallacious.”

Another submission, also using estimation, was criticized for log-transforming their data, even though the authors provided a CHI reference to explain and justify their choice.

In a different submission, reviewers reacted positively to the use of estimation:

“ALL reviewers were positive about the evaluation, and one of them even commented on the statistical method used, which they qualified to be “easier to interpret” than p Values (estimation methods are not common at CHI).”

Conversely, study reports relying on more traditional methods (i.e., null hypothesis significance testing / NHST) were occasionally criticized for being hard to understand. In one of the responses, the authors report a series of regression tables and shared a reviewer’s comment stressing that the high number of p -values in the paper was difficult to comprehend.

One author submitted a review quote where the reviewer stated that the statistics were standing too much in the foreground and a discussion of what this work meant was lack-

ing. The author commented: *"We were very thorough and one of the reviewers commented that we were rigorous but that he/she got TIRED of reading so much statistics!"*

In two cases, the authors seemed to have entirely abstained from using any inferential statistics. In response to one of these cases, a reviewer pedagogically explains that inferential statistics were needed to support the conclusion that there was no effect, especially with the presence of large statistical noise. In another case, the authors seemed to have explicitly stated known problems with NHST to justify their decision to refrain from using statistical inference and to only report summary statistics, which reviewers found unacceptable.

The review process

A number of comments had to do with the review process in general, including the role of ACs and of the rebuttal.

The Role of Associate Chairs

Twelve responses commented on AC reviews or on the role of ACs, but were overall balanced: four positive, four negative and four neutral.

Several AC reviews were praised and found extremely helpful. One respondent wrote *"1AC's comment on what needs to be addressed on rebuttal was very helpful"*, while another wrote:

"The meta-review acknowledged concerns shared across reviewers in this work, and simultaneously made a suggestion for how to find space to fix it in a dense note."

Other examples of clearly positive feedback can be found in the subsections "Encouraging Tone" and "Detailed and Vague Comments".

Other respondents however found that their AC did not fulfill their role of overseer and moderator. One author who sub-

mitted private comments wishes their AC had weighed in more in their meta-review, instead of simply summarizing reviewers' comments without considering whether these were justified. Similarly, an author commented:

"The AC should comment on all reviewers. Twice (reviews + rebuttal) my AC did not clarify on a misunderstanding from one reviewer. The reviewer did totally misunderstand the concept while all others did not. The AC did never picked this up. That's just unprofessional."

While a different respondent wrote:

"ACs should be instructed to ask the externals to revise their review when they are offending or when they write mistakes, and if the external does not respond, the AC(s) could always add a comment in the meta-review to make it explicit that that or that comment is not appropriate."

One author received no meta-review but was instead invited by the AC to do it themselves. That AC stated that reviewers raised valid points and asked the authors to read the reviews and summarize the questions. The author requested:

"More guidelines for ACs how to write a good meta review to give authors a chance in the rebuttal [...] How can we ever answer to all [questions] in a 1 1/2 page rebuttal? Without a meta review, you just pick the major ones, but that doesn't help to clear all questions of the reviewers ..."

Post-Rebuttal Comments

A relatively large number of survey responses (5 positive, 8 negative, 9 neutral) referred to post-rebuttal comments, although our survey had no specific question on this.

One author shared a long excerpt from a post-rebuttal reviewer comment as a high-quality review example. This excerpt clearly shows that the reviewer considered the rebuttal and argues why the changes described in the rebuttal would still require an additional review cycle, hence the decision remained to reject the paper. Another author writes that *"all reviewers wrote a fair answer to the rebuttal"*.

Several authors however complained that their rebuttal was mostly ignored by reviewers or that they received no post-rebuttal comments. One author suggests:

"Clearly state if there is a chance that the rebuttal could reverse the decision. If there is no way it can, then it's best to save author's time. Some reviewers do not even read the rebuttal."

On the issue of post-rebuttal comments, one author (whose paper was accepted with high scores) writes:

"We received the post-rebuttal reviews. How many of the 5 reviewers replied? TWO. How many total number of lines used per review? ONE. Did 1AC (who raised 3 issues) reply? NO. This is very disrespectful."

Even when rebuttals are considered, authors are often disappointed by how hard it is to change reviewers' mind:

"We spent a lot of time in rebuttal defending our decisions, and still only sort-of convinced the reviewer."

Another author shared a post-rebuttal comment tersely stating that the reviewer read the rebuttal, that all questions were addressed, but it did not change their mind.

Insights into the Review Process

In five survey responses, authors spontaneously mentioned that they appreciated when ACs shared information about the review process. In a meta-review comment highly appreciated by the author, the AC disclosed having played a role in reviewing the submission the year before and summarized the outcomes from last year's submission.

Another author shared a review quote pointing out that the paper was discussed at the PC meeting and the rebuttal was considered. Even though the submission was in the end rejected, the author commented:

"It was nice to know that our paper had been discussed at the PC meeting, and we know that, given our scores, this may very well have not happened. We appreciate that our work had been carefully considered by the committee."

Another author made similar comments:

"...I very much appreciate that both 1AC and 2AC made a thorough report about what happened during the PC (discussion before and during the meeting, additional AC assigned), and very well summarized what was the major point that played against acceptance. The concerns are fair, and the recommendations are helpful and encouraging."

Suggestions by Authors

One part of the survey asked respondents if they had any suggestions to improve the existing reviewing guidelines:

- Eight respondents asked for more guidance on what is expected in a rebuttal and its possible impact,
- Six respondents asked for more guidance on how to write reviews. One specifically suggests to *"maybe diversify the reviewing guidelines a bit according to the subcommittee"*,
- Five respondents asked for more open-minded reviews: *"...be more open to new ideas and new ways of doing research than asking for a boring structure that has been followed for 20 years."*
- Three respondents asked for a better way of controlling reviewer expertise, e.g.: *"Reviewers should report their expertise in individual areas (study design, statistics, engineering, ...). ACs should then assign different tasks to one or more suitable reviewers (check spelling, check for novelty, check equations, etc.) in order to make reviews more thorough."*
- Two respondents asked for more accountability, i.e., *"more Incentives for the Reviewers to actually produce high quality reviews"*, and *"means of punishing sloppy readers :)"*
- Two respondents advocated a revise-and-resubmit cycle similar to CSCW 2012 [4]: *"ACM CSCW seems to do a better job of it now, as they have a full two-pass process. I suspect this gives referees and ACs more latitude in suggesting improvements, then waiting to see how the authors handle that. Because authors respond with a paper (vs just a rebuttal), I think referees and ACs take that at-tempt more seriously. [I should note that I was initially skeptical about the ACM CSCW process]."*

Feedback on our Survey

Most respondents were very positive and were looking forward to the outcomes of this survey. Two respondents shared reservations about sharing review quotes:

"I hesitated a lot before submitting something. I suspect many people fear the "unofficial" aspect of this survey. I am myself only half comfortable reporting this, because I believe that reviewers could manage to trace back our identity. But good initiative, to try to improve the reviewing process."

"I did not feel comfortable in sharing our reviews - even as a reviewer and a metareviewer, I wouldn't want my reviews to be quoted out of context. I think we should have a better way of collecting this kind of feedback."

Our survey sparked discussions in private, in social networks and in the alt.chi review and discussion forum, revealing polarized opinions on whether it is ethically appropriate to share review quotes without explicit reviewer permission (also see discussions at [1, 9]). In our opinion, the merits of releasing such a valuable source of information likely outweigh the occasional and minor harm that may result from commenting on anonymous review quotes unfairly.

Discussion

Overall, many reviews are found to be helpful and of excellent quality. But how can we improve the others?

1. Can guidelines be improved?

Many of the the authors' concerns and proposals as captured in our survey highlight dos and don'ts that are already covered in both official and unofficial reviewing guidelines (e.g.: seek out the positive, be specific, be courteous, and avoid very short reviews [8, 7, 10, 5]). It appears that these guidelines go often ignored. How can we have reviewers read them and follow them? One possibility could be, in addition to improving guidelines, other strategies to ensure that the guidelines play some role in how reviewers assess a paper. These include structuring review forms, injecting

questions into such forms, giving more examples of using guidelines, and structuring courses/learning activities around applying them.

Other findings from our survey are less discussed and not well addressed in the official CHI reviewing guidelines. These include a desire for more transparency and honesty about the process, as well as more openness towards different methods. For instance, there are many debates about statistical practice (e.g., [3]), but multiple valid approaches exist and the choice of methods and reporting style is the author's decision. The same way authors are required to justify their use of non-standard methods, reviewers should be required to provide references when they deem these methods insufficient. However, CHI peer reviewing is not the best place to advocate one's preferred methods.

Apart from alt.chi, the review system at CHI is closed. Information flows poorly. The only way for authors to react to their reviews is through the rebuttal which usually is only read by a handful of reviewers assigned to one paper. Instead, continuous feedback should be provided to reviewers, meta reviewers, and conference chairs about authors' expectations and their experience with the process. This should be no surprise to HCI researchers. CHI does conduct surveys, it also has hard-working and dedicated volunteers, and the process is improving constantly. However, these surveys are analyzed behind closed doors and can at best influence top-down changes. While unofficial surveys such as ours cannot paint a complete picture, data can nonetheless contribute to building a richer picture and driving public discussions.

2. Guidelines for all

Reviewing guidelines are almost exclusively for *external* reviewers. Instructions specifically for meta reviewers at CHI

are rather limited (e.g., less than 300 words in the 1AC instructions for CHI 2016), despite the enormous difficulty and responsibility implied [2, 5]. Our survey respondents suggest that (at least some) ACs need more specific guidance about their roles. An excellent unofficial guide for meta reviewers has been recently written [5] and is advocated in this year's CHI reviewing guidelines. This is an excellent step forward, but will reviewers read it if its recommendations are not integrated to official guidelines?

3. *Better match expectations*

It seems that much of the distress and disappointment felt by some authors has to do with various unmet expectations about the role of meta reviewers, of rebuttals, and the amount of details in reviews. Not all expectations may be realistic. The previously quoted examples for authors appreciating more insight into the hidden parts of the review process suggest that more transparency and better communication about what can and what cannot be expected might help to match expectations from both sides. Peer review is an inherently noisy process, especially in large conferences [4]. The workload on both external reviewers and meta reviewers is extremely high, with tight deadlines. Even scientists suffer from cognitive biases, they do not easily change their mind in the face of evidence, and the relative ineffectiveness of rebuttals should not be a surprise. Why not provide a honest picture of all of this to authors?

4. *Allow reviews to circulate more freely*

Whereas guidelines are legion, actual examples of reviews and meta reviews are not. Yet a set of exemplary reviews would provide invaluable training material for reviewers and be a useful exemplification of good review practice. There are many ways a review can be poor, so we need many examples of these too. User feedback in the form of author comments on reviews are crucial too, in order to guide the

evolution of review guidelines and processes, and to facilitate open discussions on best research practices [6].

This study offers a first exploration in collecting and analyzing such user feedback. We would like to see it continued alongside the official survey for reviewers. By allowing authors to share their comments publicly and anonymously, we aim to give a voice to authors who would like to explain what they personally value in reviews, and to authors who feel their hard work has been unfairly treated and were not given the chance to express their “user feedback” in other ways. Negative emotions will necessarily surface, but feedback collected in such a way will likely be more content-rich and ultimately more productive than isolated rants from disgruntled authors near their local water cooler.

Acknowledgements

Thanks to all who took the time to participate in this survey, and to M. Kay, E. Mekler, M. R. Jakobsen, S. Reeves, and A. Tuch for providing early feedback on the survey.

References

- [1] Misc authors. 2015. Ethics of publishing received peer reviews. Online. (2015). <http://tinyurl.com/stack-reviews>
- [2] Jeffrey Bardzell. 2012. A Position on Peer Reviewing in HCI. Blog post. (27 Jan 2012). <http://goo.gl/QDsJVr>
- [3] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*, J. Robertson and M.C. Kaptein (Eds.). Springer. In press.
- [4] Jonathan Grudin. 2013. Varieties of conference experience. *The Information Society* 29, 2 (2013), 71–77.
- [5] Ken Hinckley. 2015. So You're a Program Committee Member Now: On Excellence in Reviews and Meta-Reviews and Championing Submitted Work That Has Merit. Online article. (28 Jan 2015). <http://goo.gl/wS5bk2>

- [6] Joseph Lindley and Paul Coulton. 2016. Peer review and design fiction: “Honestly, they’re not just made up”. In *CHI Extended Abstracts (Alt.CHI)*. ACM.
- [7] ACM SIGCHI. 2005. CHI Reviewing: A Guide and Examples. Online. (2005). <http://www.chi2005.org/cfp/reviewing.html>
- [8] ACM SIGCHI. 2015. Guide to Reviewing Papers and Notes. Online. (2015). <http://goo.gl/jDIDWk>
- [9] Suresh Venkatasubramanian. 2016. Who owns a review, and who’s having the conversation? Online. (2016). <http://tinyurl.com/suresh-reviews>
- [10] VGTC. 2009. Ethics Guidelines for Reviewers. Online. (11 May 2009). <http://goo.gl/bVAXAz>