Thoughts on Reviewing*

Mark Allman
International Computer Science Institute
Berkeley, CA, USA
mallman@icir.org

Categories and Subject Descriptors: K.7.0 General

General Terms: Documentation

Keywords: Referee, Publishing

1. INTRODUCTION

The July 2007 issue of CCR elicited review process horror stories. I expect that everyone has their own vast collection. I certainly do. However, I found that picking my favorite story to be like choosing my favorite offspring. Therefore, rather than focusing on a single tale of woe I have tried to extrapolate some key points from across the suboptimal reviewing I have observed. I write this essay from the perspective of an author who has years of accepts and rejects.[1] However, this note is also greatly informed by my refereeing activities over the years (on PCs, reviewing for journals, editorial boards, etc.). My intent is to make general observations in the hopes of contributing to a conversation that improves our overall review processes and ultimately helps us establish a stronger set of *community values* with regards to what we expect and appreciate in papers. While I strive for generality I do not claim the observations are unbiased or that I have closed all my open wounds in this area.

Before going any further it is important that I note several things:

- I have interacted with a large number of non-anonymous reviewers in the context of my own refereeing. I find the vast majority of those community members to be both clueful and generous with the time they donate to the community for refereeing. They strive to be careful and principled in their work. This note is not meant as an attack on the reviewing community or an attempt to lay blame on any particular person or venue. Rather, this note is about how we might improve our processes in the eyes of the authors who submit their work.

- I am *not* claiming to have written a set of glorious papers that have been unfairly rejected. I believe I have had my share of both fair and unfair rejects. Some of the papers I consider my best work have been rejected before being published and some of these are better for having been rejected.

- I am *not* claiming to have written a set of reviews that is somehow beyond reproach. I have not always adhered to the suggestions given in this note. My thinking and understanding about the review process has been and continues to be an evolution.

- I am not attempting to provide an exhaustive catalog of review problems. There are no doubt issues that I do not capture in this note. I discuss the biggest issues I have observed, but others will no doubt have their own list of "big issues". Of course, I would encourage others to address these through their own contributions to this conversation.

- Review processes will always contain some amount of randomness. While we should understand this, we should also strive to hone our review processes such that the randomness is reduced by applying a set of *shared values* to the overall process of reviewing papers.

Finally, I am writing this note about general issues in the spirit of engaging in a community-wide dialog on our review processes. I do not advertise this note as containing all the answers, but do hope to offer reasonable suggestions and to foster discussion of the topic.

2. TOUCHSTONES

First, we outline three key aspects of the review process from an author's perspective. These aspects revolve around giving authors *confidence* that their papers were treated in a balanced and professional manner.

Clarity: It is important for reviewers to remember that the review process is a black box for authors. As an author, more than anything else we appreciate understanding why a decision was made (particularly a "reject"). Even if authors do not agree with a particular reason for a rejecting a paper, it is easier to take and makes the process feel less random if the reasoning is clear and complete in the review.

Consistency: Often authors have more than one data point about a venue and therefore develop a set of expectations about the review process. These data points can come through years of submissions or parallel submissions. It is also important to remember

---

[1] A similar note could be written that outlines issues with authors from the perspective of reviewer. Perhaps I will regale you with that note at some future point. An old set of notes in this regard can be found at http://www.icir.org/mallman/plea.txt .

that while the review process may be a black box, the output of the process is not just a decision and a set of reviews, but also a program or the contents of a journal that authors will naturally look to for hints about what happened inside the black box. Reviewers should appreciate authors' expectations and strive to uphold the history of a venue. That is not to say that venues cannot change positions and approaches. But, when such changes occur they should be explicitly noted within Call For Papers and/or editorial notes.

Fairness: Ultimately, authors want to get a sense that their papers were treated fairly and the reviewers were not biased by outside factors such as who the authors are, a reviewer's personal inclination towards simulation / measurement / analytical analysis, the general topic of the paper (e.g., Yet Another Paper On ...), some small detail that the reviewer got hung up on, etc. It is extremely difficult to convey a sense of overall fairness to an author when using a blind review process. That said, it is quite easy for a reviewer to slip and introduce the perception of unfairness in a review.

3. SUGGESTIONS

This section contains a number of suggestions that reviewers could use to increase an author's understanding about and confidence in the review process. The list is not all-encompassing, but it is hoped that it hits some of the big issues that frustrate authors within our community.

3.1 Flag Good Ideas

It is much easier to find problems with papers than it is to appreciate novel ideas, hard-fought insights that come from imperfect data or the difficulty of applying some previously developed technique in a new domain. Reviews often key on the problems found in a particular paper or experimental approach. However, referees could go a long way towards giving authors confidence that their consideration is in fact well-balanced by calling out some of the good aspects of a paper in a non-glib way. Many review forms ask for "reasons to accept" a paper, but often the reasons given are superficial (e.g., "good problem area") and really do not give an author the feeling that the reviewer really tried to weight both the good and the bad of their paper.

3.2 Reviewing Ideas, Not Papers

My biggest complaint with reviewing is referees who assess the *idea* within a paper. Some referees view it as their job to reject papers because they view the idea within the paper as "unimplementable", "non-deployable", "suboptimal", "requiring hardware that does not exist" or just plain "bad" (for their own special definition of "bad"). I have been told by fellow PC members point blank that a paper is difficult to accept because of a fundamental difference with the proposed idea. I have witnessed this phenomenon often and in a wide variety of venues. Rather than focusing on the actual idea being presented, reviewers should concentrate on the execution of the paper. For instance: How well are the problem and solution motivated? How novel is the idea? How well is the idea evaluated? Are obvious and key questions addressed? How clear is the paper in explaining the idea? How well does the paper deal with the implications created by deploying the idea? How "big" is the idea?

How deep is the investigation?

All our papers are in some sense part of a collective conversation about networking. None of us submit work to have the ideas judged by (say) three random reviewers as somehow "worthy". Rather, it is the *community's job* to determine whether the ideas contained in a given paper are good, bad, need refined, can be coupled with some other idea to form some key new concept, etc. *Any* small number of community members should not act as a gate to judge ideas to determine if they are somehow "good enough" to be part of the community's conversation.

For instance:

- A paper might be written to espouse a particular concept or abstraction and so while a particular instantiation might be shown as a proof-of-concept it may not be optimal, yet the overall idea could be important and once exposed to the community better mechanisms will be designed. This phenomena is clearly evident in the vast number of "refinement" papers the community produces.

- A paper might play fast-and-loose with reality to show an upper bound. For instance, a simulation might use floating-point arithmetic when this would be far too expensive for use in real networks. Or, a scheme might assume the availability of space in a header to accommodate some new information without thinking about where the header real estate would come from. These sorts of shortcuts should not be show-stoppers for initial research. Clearly a real working system may not perform as well as the upper bound found in the paper and there will likely be additional work to be done to approximate the upper-bound behavior. However, such papers can show the promise of some idea even if it does not spell out a fully engineered solution.[2]

- Mechanisms that are today too complex to be directly implemented in (say) a router within the core network may be perfectly acceptable with tomorrow's hardware. Or, there may be neat shortcuts that another researcher will develop that will dramatically change the processing or memory footprint of the scheme without sacrificing the quality. This is more likely to happen by exposing the community to an idea, not by sending authors back to their lab.

If a particular paper suggests what a reviewer believes to be the wrong approach to the problem, one can simply view this as an opportunity to write a paper that discusses a different and perhaps better approach to further our collective conversation.

We should all show enough scholarly maturity to be able to appreciate a paper even if the underlying idea does not directly appeal to our scientific or engineering sensibilities. Further, we can help develop such thinking by showing leadership within the seminars and reading groups we take part in by drawing a distinction between appreciating a paper and appreciating the particular solution to a problem.

---

[2] In addition to reviewers not getting hung up on particular shortcuts that authors take, authors can help the situation by showing enough perspective to note these shortcuts as such and calling out the future work of nailing the specifics down and assessing the impact that would have on the results.

3.3 Be Clear and Complete

While reviews are not published documents, reviewers should not take this as a license to be less careful in their writing. A reviewer is still writing to convey a point and so should strive for clarity and completeness in their reviews. Terse reviews and/or superficial statements often do not give authors an understanding about the reviewer's reasoning, which can lead to a lack of confidence that the reviewer did their job and really tried to understand the paper at more than a superficial level. Likewise, rants about details of the paper may allow a reviewer to vent on a particular hobby-horse, but these ramblings may well fall on deaf ears as authors are likely to write off such rants more than reasoned comments.

In addition, terse reviews can hinder the process of making a decision on a paper. For instance, a short review that does not contain a reviewer's complete thinking on a paper makes it difficult for another PC member to understand the review and therefore have a discussion about the fate of the paper. Rather, what usually happens is that the PC member who wrote the short review will have to be asked to provide more details and a fuller explanation of the thinking about the paper. This is often an inefficient way to discuss a submission.

3.4 Structure

It is easy for an author to see a comment in a review and read too much into it. What is intended to be a discussion of a small issue in the paper may stick in an author's mind as the key reason a paper was rejected (or, alternatively a belief that the reviewers are off in the weeds). One way reviewers can address this problem is to structure their reviews to make it clear which points were the "show stoppers" and which were relatively minor. Often a simple list of comments or a list of issues that move chronologically through the paper are not enough to allow the authors to get a sense of which problems were key in a reviewer's thinking about the paper and which were tangential. Using headings like "big problems", "minor issues" and "nits" can go a long way towards clearly separating the issues for the authors.

In addition to authors reading too much into what are intended to be small comments, reviewers sometimes key too much on narrow pieces of a submitted paper. Receiving a review that contains much verbage about one tangential paragraph from a 14 page paper does not lend confidence that the reviewer understood or appreciated the entire paper rather than simply keying on small aspects they found troublesome. Reviewers should ultimately consider how their reviews will be received by an author and try to structure their reviews to best convey a sense of the crucial reasoning behind a decision.

3.5 Follow The Rules

Individual reviewers naturally have their own sense of what is important in a paper. Some reviewers prioritize towards big new ideas, while others appreciate deep studies regardless of whether the ultimate contribution is novel or an incremental furthering of a long line of work. Some reviewers have an internal notion of what a "workshop", "conference" or "journal" paper should look like and contain. These views often come through in reviews. However, sometimes a reviewer's world view differs from a venue's stated objective. In these cases the venue's stated objectives should win over a reviewer's natural tendencies. PC chairs and editors should be as explicit as possible to try to ensure that all reviewers are on the same page with regards to a venue's particular priorities to ensure as much consistency in assessing papers as possible.

For instance, I have observed PC members argue against papers that offer only a re-evaluation of previous results. Certainly this could well be a defensible position in a debate on the matter. However, if the venue's Call For Papers notes that such papers are acceptable then the reviewer's notion that such papers do not belong should not be considered. To consider such an objection is to run counter to the rules established in the CFP—the rules by which the author is playing. Of course, this doesn't mean that a PC member with such feelings should not try to get the CFP changed before it is issued (say, for the next year's conference).

Another place where the issue of reviewers using different "grading scales" often comes up is when determining how baked a paper needs to be for a particular venue. Some venues encourage early work and others well-executed work that is quite far along. Referees should work within the confines setup by the particular venue and, for instance, not expect too much from a six-page preliminary paper. Chairs and editors should strive to clearly outline—for both authors and reviewers—the expectations of the venue in terms of how "done" the work should be.

3.6 Papers Are Not Perfect

Reviewers should not expect papers to be perfect in all ways. This is especially true for venues that encourage early work. As an author it is frustrating to receive a review that asks for *additional* simulations, experiments or analysis seemingly without much in the way of context. We find two general categories of such comments:

- Some of these kinds of comments are seemingly suggesting experiments (etc.) that the reviewer finds to be *crucial* to supporting the claims made in the paper. In the absence of unlimited page budgets, reviewers should also address which experiments already in the paper they feel could be removed to make room for the crucial new experiments.

- Oftentimes we believe reviewers make comments about additional experiments in the spirit of trying to help authors as they continue work on the topic beyond the given submission. That is, the reviewers are engaging in a broader discussion than that of simply reviewing the paper at hand. Such comments from an uninvolved researcher are often quite valuable. However, if the review is not explicit in noting that the comments are about future work the author may not appreciate these as broader comments and might believe they are an argument about the current submission.

Reviewers are well within bounds to insist that crucial pieces are included in a paper, but should show enough perspective to understand that a paper cannot discuss *every* aspect of some topic. Therefore, it is important for reviewers to be precise as to the reason they are making comments about additional experiments. These comments can quite easily appear as a "rock fetch" to authors. In other words, no matter what the author has done and included in a paper there is always one more bit of analysis or experimentation that could be done.

3.7 Decision Process

Ultimately editors and program committees are charged with deciding whether to accept or reject a paper.[3] This decision rests not only on the reviews of the submission, but also on the discussion of the paper within the committee. Oftentimes it is difficult for an author to reconcile conflicting reviews with the ultimate decision produced. We offer three aspects that could provide authors with additional insight into the fate of their submissions.

Modify Reviews: Oftentimes PC chair encourage members to re-visit their reviews in light of the discussion about a paper that occurs after the initial reviews are written. In some cases there is some specific point that the group agrees needs to be included in the reviews and this is usually carried out. In other cases, a reviewer will adjust their score to reflect a changed opinion on the submission without changing the prose of the review. However, the general admonishment about updating the text of one's review rarely seems to move PC members to action. This can leave authors without an understanding of the critical points that led to a particular decision. Which points from which reviews were the most consequential? Which points were not given much weight? Reviewers who change their minds on points (or how much to weight those points) should be obligated to change their reviews to reflect the thinking that developed during the decision making process, as it is generally the thinking at the end of the process and not the beginning that matters most to the fate of a paper.

Summary Statements: Another way to address the problem of reviews not capturing the full discussion is for program committees or editorial boards to issue a summary of the PC discussion that better captures the key reasons for a decision and can span individual reviews. The National Science Foundation uses this process when reviewing funding proposals, which is generally a benefit for the proposers. In addition, some journal editors will try to synthesize the reviews and discussion. However, this latter practice is not consistent and even when it does happen the quality of these summaries is quite variable. We encourage PCs to consider crafting summary statements as a way to better inform authors about the discussions of their papers. A PC could enlist a student scribe to generate a first-cut of these summaries from PC meeting discussions or the workload could be spread amongst the PC members themselves.

Swap Reviews: While reviews generated within a program committee are subject to a "peer review" of sorts in that other PC members will inevitably read the reviews and discuss the recommendations and conclusions they contain. This offers (i) an incentive for reviewers to write solid reviews because they know they will have to defend them and also (ii) allows for reviews to be vetted by other community members for possible errors, omissions and oversights. While it is hard to say the degree to which (i) incentivizes good

reviews, there is no doubt that (ii) aids PCs in arriving at sound decisions by offering a double-check on reviews. It may therefore be useful to encourage the swapping of reviews when a venue does not naturally call for such sharing (e.g., in reviewing papers for a journal). This could add credibility to the review process because authors can be told that the reviews were themselves double-checked by the other referees.

3.8 Program Committee Size

One aspect that is not directly related to the information returned to an author, but can play a (sometimes large) role in the review quality and ultimate decisions is the size of a conference program committee. The difficulty of establishing a "bar" and a shared calibration of the submissions increases with the size of the PC. This, in turn, reduces the consistency of the decisions. There are some venues that receive a large number of submissions and large PCs are inevitable to deal with the deluge. On the other hand, there are smaller venues where the number of PC members and submissions are nearly the same. While this latter approach clearly spreads the load and feels inclusive, the overall quality of the program will inevitably suffer from consistency problems. On the other hand, smaller PCs add load to individual reviewers, which can in turn lead to referees spending less time on each submission in their stack—hence, lowering the overall quality of the reviews. Further, as the size of the PC decreases the breadth of expertise also may be sacrificed, possibly leading to the situation whereby referees are reviewing papers that are outside their areas.[4]

There are a large variety of considerations that go into the decision about the size and makeup of a particular PC. In this note, we simply ask that chairs consider how these decisions will impact the ultimate consistency and quality of the program when putting together a committee. Our preference is for a PC that is large enough to both handle the load and provide appropriate expertise, but no larger.

4. SUMMARY

While we believe most reviewers put forth a good faith effort to fully consider the papers they referee, we suggest a number of small changes to several aspects of the process that could aid authors understanding about how and why particular decisions were made. Steps towards clarity and transparency will ultimately give authors *confidence* that the paper selection was not a random process, but a principled and professional exercise. As noted above, I do not consider the recommendations in this note to be complete. Additional ideas in this space would certainly be useful.

Acknowledgments

---

[3]Perhaps with a slightly wider range of options open to journals (e.g., "revision requested").

[4]The impact of this depends on how general or scoped a particular workshop or conferences happens to be. In a small, scoped workshop it can be expected that PC members will generally be well-versed in the topic of all submissions, while in a more general conference a breadth of expertise is a more important consideration.