Reviewing Conference Papers

Mark Bernstein, Eastgate Systems Inc.
bernstein@eastgate.com +1.617.924.9044

In 2008, I was program chair for WikiSym, the ACM Symposium for wikis. I think I was asked to serve in this post because WikiSym is a relatively young conference and I've had a good deal of experience on related program committees[1].

When I chose the WikiSym program committee, I made a special effort to include an exceptional diversity of professional and scholarly accomplishment in order to reflect the diverse interests of the Wiki community. That meant the Program Committee included a number of business leaders and professionals who don't usually participate in academic conferences, much less on program committees. To help them, and to keep everyone on the same page, I wrote this discussion of reviewing.

In computer science, conferences are often the primary medium of scientific communication, and conference submissions are reviewed at least as stringently as journal papers. Despite the importance of the process, not much has been written on the subject.

On the Purpose of Refereeing

The primary duty of the program committee is to ensure the integrity and the reliability of the research literature. People who attend the conference, or who consult its Proceedings in the future, must be confident that the results reported here are honest, accurate, and may be relied upon.

Note that this is quite different from the duty of a commercial conference such as TED or CES, which must consider first what might best attract and entertain its customers. We would very much like to have an exciting and attractive program, but it is much more important, for us, that the research reported here be of the highest quality.

One clear consequence is that we don't particularly care who wrote a paper or where they work.

Some conferences practice *blind reviewing*, withholding the name of authors from reviewers. I find this sometimes leads to unproductive speculation and that it is helpful in some cases to know who is writing, in order to understand exactly what they mean to say.

Differences from Book and Film Reviews

Newspaper and magazine critics help people spend their time efficiently, guiding them toward the most rewarding books and films. Critics also help us understand how art functions and guide creators toward more useful and effective approaches.

But most of all, critics sell papers.

The violence and vituperation of newspaper criticism is seldom productive here.

Differences from grading papers

Conference reviewing has a certain pedagogical flavor. Many contributions are written, in whole or in part, by graduate students and postdocs. Others come from people new to the field. Part of the reviewer's role is to identify weaknesses in papers that can be remedied through additional research or better writing, and also to indicate unproductive lines of work that are unlikely to yield useful results.

But the instructor's first job is to instruct the student. The referee's job is to assemble the best available research, to show authors how it might be most effectually presented, and to help authors of unsatisfactory papers to improve them or to understand why their approach needs to change.

1 Notably the ACM Hypertext Conference on whose program committee I've served some seventeen times (including two stints as co-chair

Conflicts of Interest

Because acceptance or rejection of papers sometimes has significant personal or professional consequences, it is important that reviewers avoid conflict of interest.

What is a conflict?

You have a conflict of interest if an author of a paper assigned to you is:

- an employee of your institution or company (including subsidiaries and other faculties), or a business partner
- a collaborator or co-author of a book, chapter, or article within the past five years
- a former doctoral student, supervisor, or direct report
- engaged in a professional rivalry or feud with you
- has close personal or family ties with you

Should you find yourself having a conflict of interest on a paper to which you have been assigned, simply let me know. We have a large and capable committee, and I should have no trouble locating an alternate reviewer.

Some contributions will be submitted by members of the program committee. The committee should take care to ensure that these papers are neither favored nor disadvantaged by this circumstance.

Return papers you feel unable to read with intelligence and sympathy

In assigning reviews, I'll do my best to take into account what I know of your interests and your background. If I have erred, and you find yourself assigned to review a paper that you really cannot read, please let me know and I'll arrange for a different reader.

Occasionally, though, I may be seeking your opinion specifically because a paper is not precisely up your alley.

Obligations of a Reviewer

Sympathy

Do your best to read papers with care and sympathy. This is easy to do when the paper is good, but is still desirable even when the paper is not.

Many hours of work — in some cases, years of work — have gone into research and writing this paper. If the paper is very bad indeed, this committee may be its only audience. Do your best to approach it with a broad interest and generous spirit.

This does not argue for lenience or laxity, and certainly we cannot accept bad papers or encourage defective research.

Authors are most apt to listen to your reviews when they're thoughtful and constructive rather than caustic and dismissive.

When an unsympathetic comment is appropriate

While crushing wit and spectacular denunciation are a staple of the press critic, they are usually unhelpful in conference reviewing. An exception might be made, though, when the reviewer is certain that the author is *consistently* engaged in an erroneous course, of which the paper under review is merely the latest example. Where a mild correction might be too easily dismissed, emphasis may attract attention or provoke second thoughts.

It is usually better to attempt this correction through private correspondence or, occasionally, through public discussion — such as a conference panel or a Q&A session. But the anonymity of conference reviewing is, at times, an important safeguard of the literature.

Anonymity

Referee reports are anonymous, and in writing them it is best to be courteous but frank, candid and direct. Reviews are sent to the primary author of each paper, but the identity of the reviewer is not disclosed to authors.

A few academic reviewers prefer to sign their reviews; if you wish to do so, please include a signature in the body of the review.

In practice, the anonymity of reviews should not be too heavily relied upon. Although we take care to avoid accidental disclosure, mistakes do happen. More tellingly, authors may often be able to discern the identity of their reviewers through their interests, concerns, and styles.

Comments to the Committee

Occasionally, you may have pertinent comments to make to the committee which you do not wish to be brought to the attention of the authors. A separate section of the review form is provided for this purpose.

In general, it is better to confine your comments to those that can assist the authors in improving their work. Some useful comments to the committee might be:

I assume that equation 7, as derived in section 3.6, is correct. To be honest, I don't understand it, and I'd be more comfortable, before the paper is accepted, if someone who remembers tensor analysis could check it.

I found the paper unobjectionable but extremely dull; accepting it will cause no harm, but rejecting it would not cause much harm, either.

It might not be clear from the paper, which seems to have been adapted from some sort of proposal, that the system has actually been implemented. I have seen it, and it is actually quite promising; if the paper cannot be salvaged, a demonstration would be very rewarding.

Confidentiality

Papers submitted to the conference are confidential until they are published.

Do not use results reported in these papers in your work until the paper is published, or comes to your attention in some other way.

Rating Papers

For each paper, please provide a a rating and a reasonably complete comment. The comment is far more important than the rating, but choosing the appropriate rating will significantly assist the committee.

Scores: their limitations

Some conferences place considerable weight on numerical scores and rating. We won't do this; I find it leads to excessive emphasis on small differences.

But some ratings have a specific meaning.

The top rating is intended for very fine papers -- papers that are among the best papers published in the field, or that contain results that demand immediate publication.

The lowest rating is meant for papers that are actively pernicious. This would include those that are deceptive, duplicitous, plagiarized, or fraudulent. It would also include papers that are simply wrong.

In former times (and still in a few conferences today), acceptance was contingent upon public endorsement of the work by a member of the committee. Imagine, if you would that you would be expected to stand at the side of the stage while the paper was presented. You would rate a paper a "A" if you would be extremely proud to be associated, even in this small way, with the first presentation of this work; a "D" would mean that you'd blush to have your friends a colleagues blame you for having to sit through this.

Two additional marks are available to you to qualify your scores. An "X" represents exceptional confidence in your judgment, because the paper directly addresses topics of which you are particularly knowledgeable. A "Z" indicates that you mistrust your judgment in some degree, because aspects of the work lie outside the realms in which you feel confident.

AX: A remarkable result! One of my top students tried and failed to achieve this synthesis, and the route adopted here is remarkably clever.

AZ: The pattern language proposed in the conclusion will be influential and widely adopted; I have already begun to reconsider our firm's management practice. I admit, though, that I cannot follow the mathematics of Section 3.5 and hope that another reader will check this section with care.

DX: Prior to [Clump 96], it was widely thought that wodgets would deteriorate in the field. That there is no evidence for wodget deterioration is now clear [Dogget 99] [Egret 01], but we continue to see it used as a motivation (as here) and in grant applications.

Please note that "X" and "Z" do **not** relate to your seniority, rank, or experience at large, but only to your relationship with the specific topic of the paper. A very junior graduate student might reasonably be expected to review work on the topic of her thesis with an "X", while Turing Prize winners might frequently find themselves using the "Z".

The Narrative Is What Matters

As a practical matter, the written review is far more influential than the rating. Try to ensure that the narrative is consistent with the rating, and that both accurately reflect your opinions. Occasionally, traces of indecision or waffling make the committee's work exceptionally difficult, as people try to sort out the hidden meaning of comments.

A: What did Prof. Clump say?
B: Clump gave the paper a provisional accept. But look: almost all her comments are negative. Missing references, methodological problems, an objection to the survey technique.
C: Well, Clump always is a stickler for technique.
A: But it doesn't sound like Clump liked much of anything about the paper. And Dexter just said "reject: unimaginative and dull"
B: Which tells us nothing
A: Oh, dear.

When your reaction to a paper is mixed, it is best to explain both the perceived strengths and weaknesses – even though these might be so apparent to you as to need no mention.

Start with a Summary

The first sentence or two of your comments should summarize the key point or argument of the paper.

The summary, which might seem a tedious chore, serves two important roles. First, summaries often help explain divergent opinions. Perhaps one reader saw the paper as a theoretical study with a proof-of-concept implementation which made only a minor contribution, while another assumed that the theory was well understood but found the implementation and analysis especially useful.

Second, on rare occasions a review might be misfiled. Having a brief summary helps assure us that we're all talking about the same work.

Third, a summary not only helps the committee during the meeting; it also convinces the authors that you have read the paper with some care and sympathy. Angry authors may be inclined to dismiss your critique as incompetent or absurd; a judicious summary may help convince them to weight your arguments thoughtfully.

Major Issues

The review comment then moves to discuss any major issues that arise in reading the paper. Remember that major issues need not be disqualifying; often, they may be readily corrected. But major issues must be considered and addressed.

A very useful distinction may be drawn between papers that are actively **pernicious**, and those that are simply not particularly good.

Blunders and errors of fact

If someone relied on these results to build a bridge, might it collapse?

You discuss the behavior of meadowlarks which were observed in Colorado, Wyoming, and in South Dakota. The Dakotan birds, however, might well be Eastern Meadowlark (sturnella magna), while the other observations were doubtless Western Meadowlark (sturnella neglecta). Though visually very similar, these are two distinct species.

Errors of methodology

Please be alert for methodological errors, particularly those that could lead to misleading interpretation of the results.

You report that your students enjoyed using your system. Might they have wanted to please you? Might this be a Hawthorne effect, a response to a welcome change of pace? Might they have chosen to study with you because you are the sort of person who would create a system they would enjoy? Perhaps a study of a different population would yield new information.

Errors of interpretation

One of the most valuable services that referees can provide is simply to imagine fresh interpretations of results — and to propose ways to distinguish the alternative interpretations.

Your results are convincing and probably correct, but it seems to me just possible that, in the final insertion step, the molecule might also insert backward, forming a five-member ring with a Co-NH_2 bond instead of a six-member ring with two Co-O=C bonds. This might explain remarkable color of the product; further spectroscopy should prove conclusive.

---

multiple submissions and previously published work

Please let us know if you believe that a work has been submitted to another conference, or if it has been previously published in substantially the same form.

Failure to evaluate or reflect upon a system or feature's actual performance

Having built a new wiki or a new wiki feature, it is natural to think well of it. But research requires something more than simply designing or building a new system; we seek *knowledge and understanding,* not simply a new product.

One natural approach to addressing this question is empirical evaluation: let some number of people use the system and evaluate how effective the system is observed to be.

But this is not the only way to examine a system. We may, for example, learn from the choices people make in the course of using a system. We can sometimes learn a great deal by studying whether people choose to use a system at all, in the context of a museum installation or other public facility. In some cases, critical reflection and anecdotal experience may prove invaluable.

Many authors believe wrongly that the point of evaluation is to prove their system's strengths. Understanding the shortcomings and weaknesses of a new system is often of far greater value. Showing authors how to identify and discuss these weaknesses provides very valuable service to the literature.

Originality

WikiSym is a research conference; all papers and short papers must report new and original ideas.

Note that a thorough review of previous work makes an important and original contribution when it increases our understanding of known results, or expresses relationships that have not been properly understood or correctly interpreted.

Failure to position the work with respect to other research

In discussing original research, the author should indicate what other work is pertinent and explain how the new work relates to it.

A common misconception holds that the author should argue for the originality of her work by systematically enumerating its differences from what has been done before. This is occasionally helpful. But differences are usually easy to find in computer science and software engineering; more value can often be found by identifying analogous and related work, than by cataloging every way in which past work differs.

Failure to cite related work may anger you, especially if it's your work that is slighted. Remember that citations may be added relatively easily. In multidisciplinary areas, it can be difficult for authors to cover all of the ground, and even expert authors may overlook an essential reference. Consider how you would have regarded the paper if the appropriate citation had been added and suggest to the authors what they might cite: it may help them a great deal. Citations are seldom omitted out of malice.

Will wiki folk find the paper interesting? Does the paper belong at another conference

The WikiSym audience has broad interests, and few papers submitted to WikiSym are likely to be completely irrelevant. In some cases, though, a very narrow might find a better and more receptive audience at a more specialized conference.

In reviewing wiki papers, please take a broad view of the definition of wikis. We do not wish to limit ourselves to be the CamelCaseConference.

Opacity

Opacity might arise from several sources:

- problems with English as a second language
- mathematical or notational sophistication
- dependence on unfamiliar theory or terminology

Avoid placing too much weight on opacity; what is opaque to you might be accessible to other parts of the community, and there may be no alternative way to reach an important and useful result.

But unnecessary opacity, extraneous jargon, or incomprehensible English may render a paper useless or unread. While it is almost always preferable to treat these deficiencies in detail as minor issues, and to allow the Committee to judge whether their number precludes successful revision, overwhelming opacity may occasionally represent a research flaw that demands remedy.

Fraud

Outright fraud — false or irreproducible results, plagiarism, and other unethical practice — is seldom a concern for referees. Over the years, I have participated in program committees that have reviewed thousands of papers; in that time, I believe I have encountered one (marginally) fraudulent result and two situations where plagiarism might have been alleged. Reviewers are defenders of the literature, but in this area we are seldom called upon.

Review Articles and Practitioner Reports

The same standards of correctness, originality, scholarship and clarity that apply to research papers should also be applied to review articles and practitioner reports, but readers must bring intelligence and judgment to bear on the question as well.

A review or survey of a complex or poorly understood province of the field has considerable value, but its originality lies in interpretation and synthesis rather than the presentation of new data.

Reports of applications of systems and techniques in practice can provide invaluable information to designers and can (indeed should) inform research and engineering. It is, unfortunately, almost always easy to find fault, since practitioner reports almost necessarily report on a specific experience in specific and idiosyncratic circumstances. It is desirable for authors of experience papers to identify these idiosyncrasies, but it is the very specificity of actual experience that gives these papers their special value.

It is preferable to improve the scholarship of experience papers, even at the cost of some effort and frayed patience, rather than to reject them out of hand.

Be Resolute

In reviewing a paper, strive when possible to arrive at a clear and forthright opinion. Does the paper make a useful contribution? Does it contain inherent flaws and errors? Can its shortcomings be removed, remedied or repaired?

In the end, judge the paper as it stands. Some authors make no changes once a paper is accepted, no matter how thorough and helpful the reviews may be. When you make an accept/reject decision, consider how you would feel if the paper were published in its current state.

Most papers you review will be published somewhere, even if they are rejected here. An effective and convincing review improves the health of the research community. Rejecting a pernicious paper may simply send it elsewhere; showing the authors how to correct it and convincing them to do so is much to be preferred.

Don't spend too much time denouncing a paper with which you simply disagree (but cannot show to be wrong). Focus on the paper's correctible flaws.

Don't spend too much time in finding small flaws in a paper that suffers from such severe errors that it has no realistic chance for acceptance.

Do not be afraid to say that you don't understand the math or the theory underlying a paper you're reviewing. You will do the community a disservice if you award a high rating (or a low one) to conceal your lack of familiarity with an area. The field can be made to look foolish if one of its members makes incorrect use of sophisticated mathematical or

physical concepts. Remember, external experts can always be consulted.

When possible, be clear and forthright in championing good papers (while pointing out their correctable weaknesses) and in deploring bad ones (while taking care to indicate what strengths might be found therein and how they might be shown to advantage). Avoid the temptation to give all papers middling scores in the name of modesty or moderations.

Do not be distracted or distressed if all your assigned papers turn out to be good, or all prove bad. Your assignment represents a small sample of a large pool, and there is no reason to think it a representative or random sample.

On Minor Issues

In your reading, you will doubtless observe many points which could, or should, be improved. These are minor issues; their presence does not vitiate the work, but renders it less accessible or less useful.

Please note these minor issues as you see them. In my own reports, I prefer to discuss these in a separate section. This separation helps the Committee, and it also helps authors improve their work. My referee reports often contain far more discussion of Minor Issues than anything else.

Don't try to rewrite or line-edit the paper.

It is often useful to indicate a general rule with which the author might be unfamiliar, or may have come to neglect. It is faster for you, and more helpful to the industrious author, to point out that the paper habitually splits infinitives or omits necessary commas than to enumerate each individual error.

Style & Language

Research is ever more international in scope, and WikiSym draws interest from a vast range of disciplines. Read with sympathy and tolerance, seeking to help each author to correct errors and to express themselves clearly within the style of their discipline, and making allowance for differences in discipline and in their familiarity with the English language.

Concision

On occasion, research papers are simply too long. This most often occurs when the author is not entirely certain what is new and interesting in their work, or where they are unfamiliar with the WikiSym audience. On occasion, too, an intractable major problem may be sidestepped by drastically shortening a paper, permitting an author to present a useful result this year while pursuing further investigations of a contentious point or resolving a methodological problem. In this way, the field gains two useful publications, where a direct confrontation might lead the author to abandon the work or simply to publish it in a less discriminating venue.

Your discussion of ruby syntax (section 2) is not needed at WikiSym and can be replaced with a reference to a good manual. Section 5.1, which justified markup languages, will also be familiar to this audience. The space thus saved might be devoted to a fuller description of the novel interface animation (or to enlarging Figure 3!)

Useful Prior Work

No one can claim a truly comprehensive familiarity with all the literature that might prove useful or important to a wiki paper. An author might need, for example, to touch upon issues in software engineering, graphic design, literary theory, and graph theory. Most readers will not be intimately familiar with all these disciplines.

Citations provide credit to colleagues, starting points to students, and landmarks to scholars. To the expert reader, citation can provide exceptionally succinct information about the author's orientation and intent. Citation is especially useful in computer science, a young discipline where novelty is cheap: though not all writers understand this, it can be more valuable to draw an inexact analogy with well-known prior work than it is to claim a grand new framework.

It may often be useful to bring specific references to a writer's attention. It is useful here to indicate whether you believe a reference to be essential, and why you find it so.

Resist the temptation to insist that your own work be cited, but do bring your directly-relevant research to the writer's attention if it appears they have overlooked material that would improve the paper, or if additional citations would help clarify their meaning.

Effective Reviews

Authors are most apt to listen to your reviews when they're thoughtful and constructive rather than caustic and dismissive. Repress your savagery if you can bear to. If you can, say something positive before you launch into the critical portion of your review; authors tend to pay more attention to the review as a whole if there's something positive in it.

Your review presents your thinking both to the program committee (not all of whom may know you well) and to the authors. As the authors will not know you at all, and as they may well find any criticism you offer unwelcome, the character you present in your review will exert a great influence on its reception. Insincere praise and false camaraderie are seldom effective, but taking care to establish your knowledge, judgment, and good-will may make unwelcome advice more palatable.

A Stylistic Detail

The word "wiki" is capitalized when it refers to the original Wiki, the wiki first implemented by Ward Cunningham (http://c2.com) and its various sections and annexes, such as the Portland Pattern Repository. It is also capitalized, of course, when it is part of a proper noun (The Wikimedia Foundation, WikiSym, The International Wiki Conspiracy). When referring in general to lightweight and open collaborative systems, "wiki" is not capitalized.

The word "Web" is capitalized when it refers to the World Wide Web.

"Hypertext" is not a proper noun. A hypertext has links; "hyperlink" is not a word.

Deliberations

Every paper will be read by at least three reviewers. Please do your best to submit your reviews on time — or even a bit early if you can!

Once the reviews are in, we'll examine with particular care those papers on which readers disagreed. If reviewers clearly disagree, we'll often ask the reviewers to read each other's comments, and then to revise or extend their remarks. We might seek more readers, or solicit the opinion of experts outside the committee. The hope, here, is to find a consensus on nearly every paper, and to obtain sufficient information on those papers for which a consensus cannot be obtained.

On Saturday, June 28, some of us will meet in Porto to make final decisions on these papers. At the program committee meeting, each paper will again be examined — including papers that seems to have a consensus.

For the good papers, we ask: are there hidden flaws that readers have overlooked?

For the bad papers, we ask: is there some desirable kernel that might be salvaged for this conference by recasting or abbreviating the paper? Can the work be revised or redirected to become a strong paper next year?

For contentious papers, we ask: is the source of disagreement capable of resolution? If one reviewer has identified a clear mathematical or ideological error that others overlooked, the matter is easily settled. At times, though, the contention lies in more ambiguous grounds; in this case, we ask: (1) is there any strong reason for accepting that paper? Does it have any champions? and (2) is there any reason to think the paper actively pernicious or harmful?

Any statement of fact in a published paper might lead a graduate student, especially one working in an isolated environment, to spend several years of work pursuing a project that takes that fact as given. A special concern, then, is to address assumptions that once were widely believed but which the field now holds in doubt. Do not overlook out-of-date homilies simply because they once were accepted wisdom.

At that point, we'll select a program and pass your reports to the authors of both successful and unsuccessful submissions. Remember that the contents of papers remain confidential until publication, and committee deliberations remain confidential indefinitely.

Thanks!

Thank you once more for undertaking this work, and thank you for your patience with this long and pedantic document. This topic is not much discussed and the review process seldom described, and I thought it best to set this down explicitly, albeit at tedious length.

If you have corrections or additions to these notes, I would appreciate learning of them.

Thanks, too, to Dr. Catherine C. Marshall and Prof. Ademar Aguiar who read drafts of this paper and offered many helpful suggestion. The remaining errors are, of course, my own.

References

Alan J. Smith, "The Task of the Referee", IEEE Computer 23(4), 1990, 65-71. http://www.computer.org/portal/cms_docs_transactions/transactions/tpami/freecontent/taskoftheferee.pdf

Oscar Nierstrasz,"Identify the Champion: An Organisational Pattern Language for Programme Committees", *Pattern Languages of Program Design 4*, N. Harrison, B. Foote, H. Rohnert (Ed.), vol. 4, Addison Wesley, 2000, pp. 539-556