Low Acceptance Rates of Conference Papers Considered Harmful

Behrooz Parhami, University of California, Santa Barbara

A quantitative analysis supports the argument that very low acceptance rates of conference papers are more likely to impede publication of bold and innovative research results than to indicate the chosen papers' prestige and elite status.

Computer science and engineering (CSE) is among the handful of technical disciplines that view published proceedings papers as having equal or greater prestige than that of journal articles. As proof of their papers' elite status, some researchers cite the low acceptance rates of certain conferences and imply that this selectivity should lead to preferential treatment in an institution's research assessment. This tacit agreement about the special place of conferences in CSE has been in place since the late 1990s, when dedicated CSE conferences began to reach prominence, and has continued largely unchallenged for nearly two decades. Until recently, those outside the CSE community also unquestioningly accepted CSE conferences as the primary and preferred venues for publishing important research results.

However, this premise is being challenged on two fronts.[1] The first is that other scientific and engineering disciplines might not accept CSE's contention that journals cannot publish as rapidly as conferences. Moreover, some reports dispute the notion that quick turnaround time is essential for publishing research results in a fast-moving discipline.[2,3] Even if findings show that a shorter submission-to-publication time is preferable, other disciplines might legitimately question the assumption that all journals have long delays. Indeed, some prestigious scientific periodicals have turnaround times of days or weeks, not the months or even years typical in CSE journals.

The second front is the rampant dissatisfaction with the conference refereeing process,[4] which is characterized by time-pressured decision making; no allowance for authors to respond to evaluations; and overwhelmed program committee members,[5] who typically serve on multiple conference committees and journal editorial boards in addition to their day jobs.

To support the view that the paper-selection process for CSE conferences is flawed, I conducted a simple quantitative analysis that illustrates how conferences with low paper-acceptance rates might not be flag-bearers of high quality. Rather, such highly selective conferences tend to be more prone to the judgment errors that are unavoidable in the face of tight deadlines and overcommitted referees and program committee members. My probabilistic argument shows that when acceptance rates fall below a certain threshold, the noise in the selection process makes the quality judgments highly unreliable.

CSE research leaders are taking steps to address the problems in the paper-acceptance process and in the resulting skewed research evaluations. Institutions are training students and young researchers to become

0018-9162/16/$33.00 © 2016 IEEE

better referees,[6] looking at best practices in faculty tenure and promotion schemes,[7] reexamining publication processes,[8,9] and ensuring that evaluation systems can scale up to accommodate new modes of research-results dissemination.[10,11] I hope that this quantitative analysis will contribute to these discussions and efforts.

A QUANTITATIVE MODEL

To understand how paper-acceptance rates relate to paper quality, I developed a quantitative model that postulates n manuscripts submitted to a conference with a target acceptance rate of a, where a is some value between 0 and 1. The model assumes that submissions can be ordered according to quality, even if that order is unknown and undiscoverable in a normal selection process. An ideal (error-free) selection process accepts the top an manuscripts and rejects the bottom $(1 - a)n$ papers.

The model also assumes that an error rate, e, characterizes the conference referees: more precisely, e is the probability that a referee will recommend accepting one of the bottom $(1 - a)n$ manuscripts or will recommend rejecting one of the top an manuscripts. On the basis of my experience with both conferences and journals, an error rate of around 30 percent (0.30 probability) is not unusual. For example, it is not at all uncommon to have an "accept as is" recommendation from one of three referee reports for a manuscript that ends up being rejected. Similarly, receiving one "reject" recommendation from one of three referees is quite common for a manuscript that ends up being accepted.

The model is based on the assumption that k referees evaluate each paper and can give either an accept or a reject recommendation and that the referees' evaluations are statistically independent. Of the top an submitted manuscripts, the ideal accept batch, g, where $g = an(1 - e)^k$, will have k accept recommendations and will thus end up being accepted. Of the bottom $(1 - a)n$ submissions, f, where $f = (1 - a)ne^k$, will have k accept recommendations. Then, the total number of submissions that will garner k accept recommendations (unanimously deemed of very high quality) is h, where $h = f + g = an[(1 - e)^k + (1/a - 1)e^k]$.

It is possible for h to be less than the an target, forcing the acceptance of some submissions with mixed reviews. However, the real insight into paper-acceptance rate versus error rate comes from focusing on h. If half the h selections come from bad papers, the selection process is hardly ideal. In terms of the model, this situation requires $f = g$, or $a = 1/[1 + (1/e - 1)^k]$, which is plotted as a function of e in Figure 1 for $k = 2$, 3, and 4.

An evaluation error of approximately 32 percent with three referees per submission and an acceptance rate of 10 percent leads to the undesirable 50–50 ratio of good-to-bad manuscripts among those with accept recommendations from all referees (bold dot in Figure 1). In fact, a selection process based solely on yes or no referee edicts would be considered flawed for a 60–40 or even 70–30 ratio of good-to-bad papers. Plotting the equivalent of Figure 1 with these less extreme mixes yields curves that move upward and to the left, enlarging the region below and to the right of the curves—the region in which the selection of bad papers is much more likely.

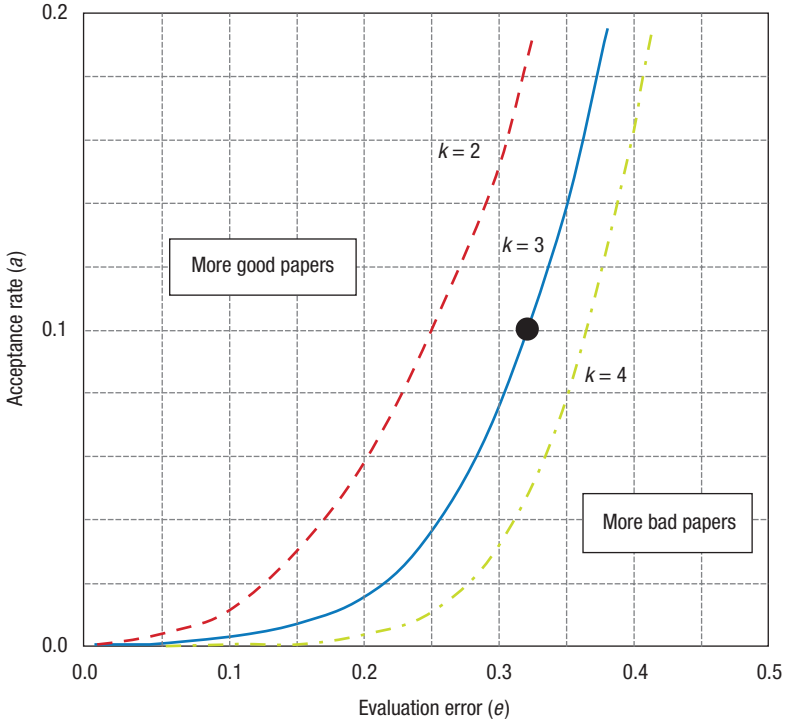

FIGURE 1. Quantitative model of paper-acceptance rate versus evaluation error with k referees per paper. In the cases shown, half of all conference-paper submissions that receive unanimous accept recommendations from two, three, and four referees would not be accepted in an ideal evaluation scheme. In the region to the right of and below these curves, a conference with a highly selective paper-acceptance rate is likely to accept more bad manuscripts than good ones. The bold dot shows a sample case in which the ratio of good-to-bad manuscripts is 50–50.

IMPROVING THE SELECTION PROCESS

This analysis applies to any research evaluation involving a single-pass, time-constrained review process in which authors lack the opportunity or have insufficient time to respond to reviewer assessments, including paper selection in conferences, journals, and research competitions.

The perils of averaging unreliable data (such as reviewer scores) are well-known in the scientific community. In his remarkable autobiographical book,[12] Richard Feynman states that even though averaging often improves estimate accuracy, averaging the estimates of random people who are not knowledgeable about a topic does not lead to improved knowledge. This observation can be extended to opinions from fairly knowledgeable people who are asked to provide numerical or binary scores quickly and without adequate time for reflection.

The pressure to produce a quality evaluation drives many of the selection-process problems my quantitative model exposes. As such, it should be the focus of improvement strategies, such as altering evaluation parameters, pairing publications, addressing blind reviews, and adjusting the basis for research rewards.

Changing evaluation parameters

Journals typically have a longer assessment time frame than conferences, as well as multiple passes of author–referee interaction. Both of these could reduce error probability and thus lead to more accurate selection.

The number of referees and referee interaction might also improve paper selection. A review process with more referees (larger k value) would be helpful, as would a face-to-face discussion about the submitted papers by the program committee. Regrettably, few conferences hold such meetings, and the meetings that are held are often poorly attended. Moreover, discussion typically begins with a blanket acceptance of submissions that received uniformly positive referee assessments and a blanket rejection of those with two or three negative evaluations. Hence, time is devoted to the papers with one negative assessment. The model's results suggest that papers should not be accepted or rejected automatically on the basis of numerical scores alone.

Publication pairing

Pairing conference proceedings publications with journal special issues allows authors of good accepted papers to prove the worth of their contributions through a secondary, often more rigorous, selection process. This dual acceptance acts as a second stamp of approval on worthy papers and weeds out those that were less deserving of initial selection. However, it cannot be retroactive and thus does not annul the appearance of bad papers in the proceedings, nor does it resurrect the good papers that were rejected because of poor selection quality.

A recent proposal under consideration by ACM is to introduce *Proceedings of the ACM*, a journal series that would publish "high-quality contributions vetted by research communities through conferences."[9] The new proceedings would systematize the current practice of publishing selected conference papers in journal special issues by effectively merging the conference and journal editorial processes. Although the proposal has some advantages, it is not universally accepted within CSE.

Addressing blind reviews

Research funding competitions, another outlet for gaining recognition for research programs, also undergo time-constrained reviews and typically have low proposal-acceptance rates. This trend is due primarily to the funding entity's prestige or budgetary shortfalls. An added complication is the undesirability of blind reviews, given that the researcher's qualifications and track record are legitimate criteria for assessing a research proposal.

Although a researcher's name or institutional prestige can lead to the acceptance of some good manuscripts more or less automatically, averaging submission quality scores is still a danger. Consider a conference that accepts 15 percent of the papers submitted. If 10 percent of the submissions come from recognized researchers with unquestionable research quality, then the acceptance rate for the remaining 90 percent of the submissions is only 5.6 percent [(0.15–0.10)/0.90]. This significantly lower acceptance rate for the bulk of submissions amplifies any judgment errors. Such a conference will publish a mix of good work from the recognized researchers and some more or less randomly chosen works from the rest of the research community. With reputation being the key selection factor, new researchers have a lower chance of acceptance, which can have potentially devastating consequences.

Rewards adjustment

Referee reports produced under extreme time pressure or a refereeing

commitment that is a fraction of a heavy workload undermine quality paper selection. Part of the problem is that conscientious refereeing activities are not well rewarded. Our institutions must work on adjusting current reward systems and emphasizing the ethics of research assessment in CSE graduate programs.

This article's title mirrors the title of Edsger Dijkstra's "Go To Statement Considered Harmful," a short letter to the editor that exposed the negative impacts of go-to statements on a program's readability and ease of software debugging or maintenance.[13] In a similar vein, this article's intent is to expose how conferences that reject an overwhelming majority of submissions might harm scholarly research in CSE and other technical fields.

My quantitative model shows that the lower the paper-acceptance rate, the lower the error rate must be to maintain reasonable quality—a requirement that is at odds with quick decisions. Given that the error rate is pretty much dictated by referee turnaround time and load, lowering the acceptance rate has the undesirable effect of worsening the ratio of good-to-bad manuscripts selected for publication. The current manuscript selection process might not be as unreliable as my model suggests, but only because time-pressured referees tend to base their recommendations on an author's or a research group's reputation rather than on an independent quality assessment. This trend penalizes new researchers and discourages the publication of bold new ideas.

Although it has become commonplace in CSE for authors of papers published in a proceedings with a highly selective paper-acceptance rate to wear their acceptance as a badge of honor, these very low acceptance rates are detrimental to the CSE profession as a whole. In retrospect, using conferences as the primary publication venues in CSE was the wrong solution to the very real problem of slow turnaround for research publications. A better path would have been to make journals more responsive to the demands of a fast-moving field. Fortunately, a movement toward this goal seems to be afoot. C

ABOUT THE AUTHOR

BEHROOZ PARHAMI is a professor in the Department of Electrical and Computer Engineering and a faculty participant in the computer engineering program at the University of California, Santa Barbara. His research interests include computer arithmetic, parallel processing, and dependable computing. Parhami received a PhD in computer science from the University of California, Los Angeles. He is an IEEE Life Fellow and an associate editor of *IEEE Transactions on Sustainable Computing*. Contact him at parhami@ece.ucsb.edu or www.ece.ucsb.edu/~parhami.

REFERENCES

1. M.Y. Vardi, "Conferences vs. Journals in Computing Research," *Comm. ACM*, vol. 52, no. 5, 2009, p. 5.
2. L. Fortnow, "Viewpoint: Time for Computer Science to Grow Up," *Comm. ACM*, vol. 52, no. 8, 2009, pp. 33–35.
3. J.Y. Halpern and D.C. Parkes, "Journals for Certification, Conferences for Rapid Dissemination," *Comm. ACM*, vol. 54, no. 8, 2011, pp. 36–38.
4. H.F. Korth et al., "Paper and Proposal Reviews: Is the Process Flawed?" *ACM SIGMOD Record*, vol. 37, no. 3, 2008, pp. 36–39.
5. K. Birman and F.B. Schneider, "Program Committee Overload in Systems," *Comm. ACM*, vol. 52, no. 5, 2009, pp. 34–37.
6. A.J. Smith, "The Task of the Referee," *Computer*, vol. 23, no. 4, 1990, pp. 65–73.
7. D. Patterson, L. Snyder, and J. Ullman, "Evaluating Computer Scientists and Engineers for Promotion and Tenure," *Computing Research News*, 1999; www.cra.org/uploads/documents/resources/bpmemos/tenure_review.pdf.
8. D.S. Wallach, *Rebooting the CS Publication Process*, tech. report, Rice Univ., Aug. 2010; www.cs.rice.edu/~dwallach/pub/reboot-2010-06-14.pdf.
9. J.A. Konstan and J.W. Davidson, "Should Conferences Meet Journals and Where?: A Proposal for 'PACM,'" *Comm. ACM*, vol. 58, no. 9, 2015, p. 5.
10. J. Crowcroft, S. Keshav, and N. McKeown, "Viewpoint: Scaling the Academic Publication Process to Internet Scale," *Comm. ACM*, vol. 52, no. 1, 2009, pp. 27–30.
11. D. Terry, "Publish Now, Judge Later," *Comm. ACM*, vol. 57, no. 1, 2014, pp. 44–46.
12. R.P. Feynman, *Classic Feynman: All the Adventures of a Curious Character*, R. Leighton, ed., W.W. Norton & Co., 2006.
13. E. Dijkstra, "Go To Statement Considered Harmful," *Comm. ACM*, vol. 11, no. 3, 1968, pp. 147–148.