Varieties of Conference Experience

Jonathan Grudin
Microsoft Research
One Microsoft Way
Redmond, WA 98052

The dilemma

Two primary goals of a research field are (i) communicate and preserve knowledge, and (ii) build and maintain a professional community. A range of tools are at its disposal: books, journals, newsletters, lectures, exhibitions, person-to-person communication, magazine and newspaper articles, workshops, and conferences. Each field is marked by a distinct mix and weighting of these venues. To showcase quality work, humanities emphasize books whereas sciences typically look to journals. For community-building and member support, conferences and workshops support networking and often provide community members with feedback on their work in progress.

When everything proceeds smoothly, the complexity of the interdependencies and interacting forces can be difficult to appreciate. By analogy, the internal mechanisms of a healthy running motor are difficult to deduce. Perturb the mechanism and the vibrations and sounds of a motor about to fail can be illuminating.

Readers of recent issues of Communications of the ACM have encountered vibrations and sounds that signal a crisis in computer science. Few issues do not include someone shouting that the motor is about to fly apart. My voice was neither the first nor the last when I recently suggested a source of the perturbation (Grudin, 2011). A prescription is not easily found, however, and the problem may be spreading—including to information science.
In research, as in an organization, introducing a new communication or collaboration technology disrupts the status quo, the existing balance of communication channels, information repositories, and loci of authority and responsibility. If the new technology takes hold, a presumably more effective ecology eventually emerges. But there can be unintended consequences, and not every mutation proves to be adaptive.

My hypothesis is that word processing and the Web, in the context of the professional culture of North American computer science in the 1980s, led to an odd change: Major conferences became the venue for demonstrating and preserving high quality computer science research in the US. In the process, these conferences became highly selective. This disrupted community maintenance, the conferences’ original raison d’être. Serving two purposes has stressed the engine.

**Computer science: Travails of the pioneers**

In the 1980s, word processing was available to most computer scientists. It enabled the relatively inexpensive production of decent proceedings prior to a conference. Then—and this is critical—ACM, a non-profit professional organization, stockpiled bound proceedings and made them available effectively in perpetuity at a very low price. Access was further simplified a decade later with the appearance of ACM’s digital library, freely available by site license at most research institutions.

With peer-reviewed proceedings widely accessible, conferences suddenly filled the traditional role of journals: archiving and communicating results broadly. Emphasizing quality given its archival status was a natural next step. Conference reviewing increased in rigor; it is quicker albeit less comprehensive than journal reviewing. Conference reviewing did not support revision cycles, but rejected papers could be revised and submitted to the next conference or to a different one. As paper proceedings gave way to digital proceedings distributed on flash drives and online, conference length limits relaxed, and ACM now permits digital appendices. Further blurring the distinction, journals increasingly expedite reviewing
and constrain article length. Fast and concise communication has greater value in our information-rich world.

Despite some claims to the contrary, the shift to conferences in the US was not due to the pace of computer science research. In Europe and Asia, conferences did not achieve the same level of prominence. No professional organization stepped in to archive proceedings, so research reported at European and Asian conferences still had to be progressed to journal publication to be widely accessed.

It seemed all goodness, as some of my colleagues are wont to say. But indirect effects gradually emerged. The social psychologist Joseph McGrath observed that groups engage in activities in support of three functions: production, member support, and healthy group dynamics. All are crucial, but we tend to focus on the first and trust our natural instincts as social animals to take care of the others. We had evolved venues supporting all three: journals to showcase our research results; workshops to provide members with visibility and feedback; conferences to build a sense of community with shared purpose, congenial sites for working out organizational dynamics and polishing work. Casting conferences into a quality production role created conflicts with the community maintenance.

In a breathtakingly short time, some conferences that had firmly rejected quotas (“we accept all publishable papers”) were firmly limiting acceptances to 25% or less. At first blush, the increased selectivity seemed fine: Rejected work could after all be polished and resubmitted to the next conference, not unlike a journal’s revise-and-resubmit judgment. But over time, problems became evident.

1. High rejection rates spawn specialized spin-off conferences. The original conference publishes only a small sample of the research in any sub-topic. Many salvageable rejected submissions are available. A journal revise-and-resubmit locks authors into that journal; not so with conference rejections.
2. The specialized spin-off conferences absorb organizational energy, travel budgets, and loyalty. Authors may identify more with a specialized community that accepts more of their work. Nor are the papers necessarily of lower quality after being revised for resubmission (see 5 below).

3. Unlike large conferences in other fields, highly selective conferences lose the “must-attend to find out what is happening” aura. Faculty may send students but not attend regularly, diminishing the sense of community at the event. When the point of submitting was to attend, attendance by authors wasn’t enforced; today it often is.

4. People simply do not feel warmth and acceptance when conferences reject their work.

5. Consciously or unconsciously, people recognize that selection is often arbitrary. Choices among submissions of equivalent value depend on vicissitudes of reviewer assignment (Anderson, 2009) and perhaps the time of day a review is written (The science of justice, 2011), or where in the program committee meeting schedule a paper is considered (Nathalie Riche and Philip Wadler, personal communications). One colleague, upset about four rejections, fumed, “this is the last year I only submit four papers!” If quality was all-important, focusing on improving one or two would make more sense. But treating it as a lottery and buying more tickets worked well for him the following year.

6. The zero-sum game that quotas produce has intensely corrosive effects on group dynamics. One sub-group after another claims that a majority bias against it; my objective measures of keywords selected by authors, in cases I have checked, do not support the claims of persecution. Another consequence is the occasional gaming of the process. High selectivity is widely believed to favor incremental over original work; it unquestionably impedes practitioners or researchers from other fields who would like to present papers.
Through the 1980s and 1990s this was a US computer science phenomenon. Elsewhere, conferences remained places where colleagues contributed to polishing fairly complete work not yet in final form. This is changing.

**Avoiding the fire but not the frying pan?**

Through trial and error, my conferences found that acceptance rates between 20% and 25% were as low as we could go without risking rapid collapse. With lower rates, submissions and attendance declined. At 20%-25%, attendance was fairly steady and submissions eventually rose for major conferences. Conferences with a powerful attractor other than research can adopt even lower rates, such as SIGGRAPH’s exhibitions and ICIS as a place for employers to meet job seekers.

An acceptance just high enough to avoid sudden death may produce a chronic illness. The commentaries in *Communications of the ACM*, which are listed in Grudin (2012), describe disease symptoms but have not mobilized a response. Like the proverbial frog boiled by a slowly rising temperature, we may feel uneasy but not uneasy enough to jump.

Gloria Mark (personal communication) analyzed ACM conferences and reports that a rise in the rejection rate in year N is correlated with a reduction in submissions in year N+1. Attendance at many computer science conferences peaked even as the ranks of academics and professionals grew; in some cases it declined appreciably. Membership in associated societies has fallen. Table 1 shows participation in the largest ACM Special Interest Groups between 1990 and 2010. The continuous decline was sharpest in the early 1990s, prior to significant Web activity but when conferences were driving down acceptance rates. A further 6% decline occurred in 2011, with student membership declining 8%. New SIGS on trendy topics appear, but none have thrived.
There are undoubtedly other factors contributing to the health of a field, but the conflicted role of conferences is strongly implicated in the ills afflicting the computer science community.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PLAN</td>
<td>12,335</td>
<td>6,574</td>
<td>4,362</td>
<td>2,765</td>
<td>2,323</td>
</tr>
<tr>
<td>GRAPH</td>
<td>11,811</td>
<td>6,986</td>
<td>6,298</td>
<td>7,902</td>
<td>7,216</td>
</tr>
<tr>
<td>SOFT</td>
<td>10,824</td>
<td>6,389</td>
<td>3,313</td>
<td>2,916</td>
<td>2,489</td>
</tr>
<tr>
<td>ART</td>
<td>8,955</td>
<td>3,707</td>
<td>1,917</td>
<td>1,559</td>
<td>1,123</td>
</tr>
<tr>
<td>OPS</td>
<td>6,801</td>
<td>4,154</td>
<td>2,356</td>
<td>1,793</td>
<td>1,828</td>
</tr>
<tr>
<td>CHI</td>
<td>5,023</td>
<td>5,186</td>
<td>4,950</td>
<td>4,409</td>
<td>4,415</td>
</tr>
<tr>
<td>ARCH</td>
<td>4,685</td>
<td>3,035</td>
<td>1,730</td>
<td>1,454</td>
<td>1,384</td>
</tr>
<tr>
<td>ADA</td>
<td>4,089</td>
<td>1,867</td>
<td>685</td>
<td>391</td>
<td>292</td>
</tr>
<tr>
<td>MOD</td>
<td>3,952</td>
<td>2,959</td>
<td>2,941</td>
<td>2,317</td>
<td>1,922</td>
</tr>
<tr>
<td>MIS</td>
<td>3,082</td>
<td>1,442</td>
<td>755</td>
<td>636</td>
<td>497</td>
</tr>
<tr>
<td>(all 30+)</td>
<td>103,489</td>
<td>65,698</td>
<td>47,042</td>
<td>44,600</td>
<td>41,008</td>
</tr>
</tbody>
</table>

Table 1. Membership in the top 10 Special Interest Groups of 1990. Today only two of 37 SIGs have 3000 members.
**Information science**

Although largely restricted to the United States for two decades, recognition of selective conference proceedings has spread. The catalyst, accessible proceedings, no longer requires the resources ACM provided in the 1980s and 1990s. European universities and funding agencies now give more weight to selective computer science conferences.

Information science is trying this approach. The iConference and ASIST conferences now archive their proceedings in the ACM Digital Library. Both have lowered their acceptance rates; a few years ago they were more inclusive, with acceptances ranging into the 60% or higher; most recently both were in the 30% range. If this pattern is maintained, indirect effects on participation could emerge. The hope is that selectivity will increase stature and attract participation, but as noted the pattern for major computer science conferences is mixed: more submissions, possibly less varied submissions, more stress for reviewers, and declining attendance.

**Assessing the present**

The problem cited most often is reviewer burn-out. Conference proliferation is a factor, dividing reviewer attention and loyalty. Authors who adopt a shotgun submission approach to overcome the role of chance in reviewer assignment and acceptance add to the workload. High rejection rates lead to resubmission to two, three, or more conferences, requiring new reviewers each time. Another factor is a steady increase in the number of reviewers assigned to papers at some major conferences, perhaps a result of the conscious or unconscious recognition of the randomness in the process. This recognition adds to the stress felt by reviewers who must find grounds for rejecting papers that they sense do not differ in quality from many that are accepted.
Some recent efforts to improve the review process increase the work for reviewers: Ranking papers rather than rating them, publicly posting reviews of accepted papers, and permitting author rebuttals of reviews prior to final decisions. The first two have generally been tried only once by a given conference. If rebuttals affect outcomes, the art of writing one is a new skill for authors to master, favoring insiders such as past program committee members and their students.

High selectivity is alleged to favor incremental over original work. Authors can recycle the justification, literature review, methods section, and analysis techniques from the published work that is being built upon, whereas highly original work is likely to include a lapse that alienates a reviewer. Researchers can methodically use this to game the system by finding an old published study that can be replicated in a one-off study with new technology that is guaranteed to perform better, even when there is negligible chance of the findings being used because the field has standardized in that area and moved on.

Researchers are now trained to produce 20-minute presentations and 10-page papers. Developing a longer synthesis or talk can stymie even a strong student. In the past, we could develop our ideas through a series of formal but ephemeral presentations. Students today often must jump from very rough drafts discussed informally to polished work headed for their permanent oeuvre. Rather than safe harbors in which to get feedback on research that is largely complete but not yet polished, selective conferences are directly exposed to the stormy sea that is the Web. An added complication is that once work is archived online, self-plagiarism concerns often prevent authors from publishing refinements.

Various corrective measures are proposed:

i) Return to the good old days when journals were for archiving and conferences for conferring

ii) Tinker with existing conferences to improve them or create conference-journal hybrids

iii) Envision a very different future and take steps toward preparing for or realizing it
Actually, I’m unaware of anyone proposing (iii). My conclusion will include thoughts along this line. (i) is often endorsed but will be difficult to implement. Many senior researchers sent their journals to a shredder years ago as they built careers publishing their best work in conferences. They will resist retroactive devaluation of those conferences. However, ACM took a step to keep this option open by prohibiting incremental changes that could further depreciate journals. This action is described in the following summary of recent experiments in computer science conference organization.

**Conference-journal hybrids**

As selectivity increases, the temptation mounts to declare that a conference proceedings is an issue of a journal. The most prominent of several conferences doing this is ACM SIGGRAPH; since 2002 its proceedings is the fall issue of *ACM Transactions on Graphics*. Another is Infovis, with proceedings becoming an issue of *IEEE Transactions on Visualization and Computer Graphics*. To control this blurring of conferences and journals, ACM is considering a policy that proceedings can only be designated ACM journal issues if an open-ended review process precedes the conference. (SIGGRAPH may be allowed an exception.)

A conference-journal that follows this model is *Proceedings of the Very Large Data Bases Endowment (PVLDB)*, initiated in 2008. Papers accepted by May become the sole content of the annual fall conference. The ACM Computer-Human Interaction (CHI) and Intelligent User Interfaces (IUI) conferences include presentations of articles published in previous *Transactions on Computer-Human Interaction* and *Transactions on Intelligent Interactive Systems*, respectively. CHI and IUI also include papers submitted to and reviewed for the conference, which are not designated journal articles.

Whether to justify elevating conference papers to journal status or simply to improve them, some conferences include a mentoring or shepherding step, for some or all papers. SIGGRAPH and Infovis do this. Shepherding is an attractive idea, albeit requiring more work in a compressed time for exhausted
reviewers. It may encourage authors to revise final versions more than they would have, but a shepherd is often an advocate who pushed the paper through a tough selection process. The SIGGRAPH 2011 experience is typical: 100% of the provisionally accepted papers ended up being published.

Some conferences have inserted a real revision cycle. Aspect-Oriented Software Development (AOSD) has three submission deadlines. For the 2012 conference, authors could submit in March, July, or September of 2011. The March and July review processes yielded accept, reject, or revise & resubmit decisions, with revisions considered in subsequent review phases.

The Computer Supported Cooperative Work conference has adopted a process resembling a rapid journal special issue. I was co-program chair of CSCW 2012, which extended the review period two month to accommodate a full revision cycle. CSCW usually accepts 20% to 25% of submissions. In the first round, 45% of the 415 CSCW 2012 submissions were rejected. The others had a month to produce a revision and document outlining their changes. Those that normally would have been accepted ‘as is’ were improved as well as the others, many very substantially. In the second round, 28% of the remaining submissions were rejected, yielding an overall 40% acceptance rate. The consensus was that quality and quantity increased. Analysis of the previous year’s data enabled us to focus reviewing on papers that needed it; less was devoted to the bottom 45% and top 10%. As a result, reviewing per paper decreased despite the revision cycle. By accepting 40% rather than 20%-25%, we reduced subsequent resubmissions, sparing the community additional reviewing effort.

Freed from the zero-sum-game imposed by an acceptance quota, reviewer teams could help an author without feeling it disadvantaged other authors. Some program committee members remarked that for the first time, the review process left them energized rather than exhausted. This approach has been adopted by other conferences, but some in the community are unhappy with the higher acceptance rate. It may take time to assess our conviction that quality was enhanced. Despite this relatively
successful outcome, living with the process for over a year led me to conclude that large conferences face an insurmountable problem.

The large conference dilemma

Thomas Anderson (2009) makes a convincing case, supported by my observations, that a conference attracts some strong submissions, some weak ones, and a large number in the middle that vary little in quality. A selective conference accepts 10% that most would regard as strong, dismisses about half that attract no positive reviews, and then in essence conducts a lottery to select about a third of the remaining contenders. The lottery may spur authors to aim for the top 10%, but may more often encourage shotgun submissions: Buy a lot of lottery tickets to increase the odds of getting a winner.

In the range of negligible quality difference, outcomes rest with reviewer assignments. Reviewer variability is inescapable. Not one of our 415 submissions received a top ‘5’ from each of the first three reviewers. (A handful received three ‘1’s.) Every paper is at risk of being assigned to uncongenial reviewers. We made a strong effort to match appropriate reviewers to submissions. We carefully devised a set of topic areas for authors and reviewers to label their work and interests. Program committee members could bid on submissions based on title and abstract. We looked at different ways to use topic labels and bids to make assignments. But my conclusion was that especially when hundreds of assignments must be made in a short time, reviewing will never be equitable.

Reviewers and authors who share interest in a topic may differ in preferred methods, analytic approaches, or theoretical dispositions. Reviewers who agree on the virtues and flaws of a paper come down very differently: some are Santa Clauses, others are systematically harsh. One of our most eminent reviewers said with evident pride, “I know that I am a hard ass compared with other reviewers.” Authors assigned to him did not fare well. Some conferences use statistical normalization to address this. Any data could be useful, but in my analysis, the benefit is limited. Eventually deliberations
focus on the text of the written reviews. Our data were clear: To be accepted, an advocate is not sufficient, but is necessary. Statistics that show that a paper had the bad luck to be assigned chronically negative reviewers does not produce an advocate or remove the harsh reviews. Harsh reviewers often document their case more carefully than positive reviewers who assume a paper will win easy acceptance and focus on possible improvements (that is, point out flaws).

Adding to the variability, reviewers may be influenced by the time since they last ate when they review (The science of justice, 2011). We found some strong negative as well as positive swings between successive reviews of a paper and its revision.

The good news is that if contested submissions are of indistinguishable quality, the conference quality is unaffected by randomness. The bad news is that if we reject many in order to claim that we are discriminating, we may breed dissatisfaction. Authors of rejected papers see accepted papers that are no better (and strike the rejected author as being objectively worse). We are claiming to provide more than the fairness of a lottery. After 25 years of participating in and analyzing these processes, I do not think a large selective conference can be very fair. It can be accepted, with grumbling, but the consequences are bad for community health: disaffected, rejected authors and stressed reviewers who are mostly dedicated, intelligent, and fair people, yet who realize at some level that their efforts are falling short of the ideal.

A hypothetical conference

The binary nature of review outcomes is at the heart of the problem. Quality falls along a continuous curve and there is noise in measuring it, but the decision is sharply discontinuous. Conferences that accept virtually everything exist and can be good for community-building—a leading neuroscientist told me that the annual Neuroscience conference, at which 15,000 papers were given, was a must-attend
conference because “it is where you find out what is happening.” But this is not a viable path for fields that have come to rely on conferences for quality. The following proposal seeks a middle ground.

Consider an archival conference in which reviewers produce a 1-5 overall rating. A paper that draws an average of 4 or better has good prospects, less than 3 is not good, and 3.25 or 3.33 is in the active discussion range. In this hypothetical conference, submissions reviewed as usual. Authors receive the reviews and are given a choice: (i) Withdraw the paper, perhaps to revise and submit elsewhere; or (ii) Have the paper presented and archived, with the rating average appended to the title and used with the topic to place it in a session at the conference. Authors of a 2.0 paper could leave it in, but colleagues, tenure committees, students, family and friends could forever see that the authors were happy publishing work so rated by their colleagues. At the conference, their session might attract just the authors of the 2.0 papers being presented plus a few people interested enough in the topic to see unpolished work—to find out what is happening.

Conference attendees and proceedings readers could easily focus on highly-rated work. An author with unfortunate reviewer assignments could decide to stay in; if a revision cycle is included, they could try to raise their average. As was common in my conferences when they more inclusive, practitioners who wish to present less-polished work for the benefit of other practitioners could do so, obtaining travel funds and feeling part of the community. Tenure committees too busy to form independent judgments of work quality could focus on how the work is received and cited rather than just the conference name.

**Conclusion: Is there really a problem?**

In the mid-1990s, I concluded that highly selective conferences, where reviewing focuses on flaws or nits to justify rejection, were damaging the field. I shifted to reviewing for ~50% acceptance conferences, where the task is to identify what is interesting in a submission. But the prevailing winds blow in a
different direction. Many people feel that low acceptance rate signals quality. Students and junior faculty in my field believe that acceptance rates are critical in appointments and promotions.

These are questionable claims. A journal that nurtures papers along will raise the acceptance rate and the quality. Academic decisions may rely more on letter and personal impressions. But there are forces that cannot be ignored. When conference proceedings are made accessible online, and there is increasingly less reason not to do so, it is natural to gravitate toward expecting higher quality. Quality can be respected without harming community, as journals have long demonstrated. But journals alone do not build and maintain community, nor are they effective bridges to practitioners and other communities.

Most of the innovations that I described are pushing conferences to be more journal-like.

Simultaneously, journals are pressed to be more conference-like, publishing shorter, more accessible articles with quicker turn-around. These changes are consistent with the evolving demands of information management. Many scholars no longer have the luxury of immersing at length in another person’s work. So much information is available on so many topics that today’s crucial skills are browsing, skimming, and synthesizing material, not analyzing one work in depth. People often learn more efficiently in artfully arranged shorter bursts. It is inefficient to deliver information in short bursts in classrooms, but it can be done online. When print enabled journals, it left the basic form of the book and the university intact. Digital technologies are transforming them all.

Reflect again on McGrath’s three concerns: production, member support, and group well-being. The production of research will continue. Less careful attention is paid to the other functions.

Member support. Where, in a world where everything is digitally recorded, will young researchers find safe places to try out unpolished ideas? Thirty years ago, conferences were such places. Workshops may serve, until they too are archived, but the proliferation of conferences and increasing pressure to
publish or perish undermines less formal venues. Perhaps young researchers will grow up more quickly or be more open to making their unpolished work publicly visible. They grew up with digital cameras and camcorders capturing their casual activity. If their Facebook history accompanies them through their lives, why should early stage ideas embarrass them?

Community. The shifting role of conferences disperses the research communities that once congregated around them. The reduced sense of belonging to one community threatens institutions built on voluntarism, notably reviewing. This is a major challenge for computer science, information science, and other fields that are drawn along the same trajectory.

When a solution emerges, it may not involve conferences as we know them. It seems likely that communities will move online. After decades of research that never translated into use, technologies for immersive remote video and computer interaction are being realized in the gaming world. Pressure to cut down on travel and carbon footprint adds to the impetus. Online communities may be flexible and short-lived, in keeping with their times. Erosion of today’s research communities could clear the way for a major transformation that many sense is coming.

References


