The *Communications* Web site, <http://cacm.acm.org>, features more than a dozen bloggers in the BLOG@CACM community. In each issue of *Communications*, we'll publish selected posts or excerpts.

Follow us on Twitter at <http://twitter.com/blogCACM>

DOI:10.1145/2732417

<http://cacm.acm.org/blogs/blog-cacm>

The Arbitrariness of Reviews, and Advice for School Administrators

John Langford examines the results of the NIPS experiment, while Mark Guzdial considers the role of class size in teaching computer science.

John Langford
“The NIPS Experiment”

<http://bit.ly/1wZeUDK>
January 8, 2015

Corinna Cortes (<http://bit.ly/18I9RTK>) and Neil Lawrence (<http://bit.ly/1zy3Hjs>) ran the NIPS experiment (<http://bit.ly/1HNbXRT>), in which one-tenth of papers submitted to the Neural Information Processing Systems Foundation (NIPS, <http://nips.cc/>) went through the NIPS review process twice, and the accept/reject decision was compared. This was a great experiment, so kudos to NIPS for being willing to do it and to Corinna and Neil for doing it.

The 26% disagreement rate presented at the NIPS conference (<http://bit.ly/18Iaj4r>) understates the meaning in my opinion, given the 22% acceptance rate. The immediate implication is that one-half to two-thirds of papers accepted at NIPS would have been rejected if reviewed a second time. For analysis details and discussion about that, see <http://bit.ly/1uRCqCF>.

Let us give P (reject in 2nd review | accept 1st review) a name: arbitrariness. For NIPS 2014, arbitrariness was ~60%. Given such a stark number, the primary question is “what does it mean?”

Does it mean there is no signal in the accept/reject decision? Clearly not—a purely random decision would have arbitrariness of ~78%. It is, however, notable that 60% is closer to 78% than 0%.

Does it mean the NIPS accept/reject decision is unfair? Not necessarily. If a pure random number generator made the accept/reject decision, it would be ‘fair’ in the same sense that a lottery is fair, and have an arbitrariness of ~78%.

Does it mean the NIPS accept/reject decision could be unfair? The numbers make no judgment here. It is a natural fallacy to imagine random judgments derived from people imply unfairness, so I would encourage people to withhold judgment on this question for now.

Is arbitrariness of 0% the goal? Achieving 0% arbitrariness is easy: choose all papers with an md5sum that ends in 00 (in binary). Clearly,

there is more to be desired from a reviewing process.

Perhaps this means we should decrease the acceptance rate? Maybe, but this makes sense only if you believe arbitrariness is good, as it will almost surely increase the arbitrariness. In the extreme case where only one paper is accepted, the odds of it being rejected on re-review are near 100%.

Perhaps this means we should increase the acceptance rate? If all papers submitted were accepted, the arbitrariness would be 0, but as mentioned earlier, arbitrariness of 0 is not the goal.

Perhaps this means NIPS is a broad conference with substantial disagreement by reviewers (and attendees) about what is important? Maybe. This seems plausible to me, given anecdotal personal experience. Perhaps small, highly focused conferences have a smaller arbitrariness?

Perhaps this means researchers submit to an arbitrary process for historical reasons? The arbitrariness is clear, the reason less so. A mostly arbitrary review process may be helpful in that it gives authors a painful-but-useful opportunity to debug easy ways to misinterpret their work. It may also be helpful in that it rejects the bottom 20% of papers that are actively wrong, and hence harmful to the process of developing knowledge. These reasons are not confirmed, of course.

Is it possible to do better? I believe the answer is “yes,” but it should be understood as a fundamentally difficult problem. Every program chair who cares tries to tweak the reviewing process to be better, and there have been

many smart program chairs that tried hard. Why is it not better? There are strong nonvisible constraints on the reviewers' time and attention.

What does it mean? In the end, I think it means two things of real importance.

1. The result of the process is mostly arbitrary. As an author, I found rejects of good papers hard to swallow, especially when reviews were nonsensical. Learning to accept the process has a strong element of arbitrariness helped me deal with that. Now there is proof, so new authors need not be so discouraged.

2. The Conference Management Toolkit (<http://bit.ly/16n3WCL>) has a tool to measure arbitrariness that can be used by other conferences. Joelle Pineau and I changed ICML 2012 (<http://bit.ly/1wZiZaW>) in various ways. Many of these appeared beneficial and some stuck, but others did not. In the long run, it is things that stick that matter. Being able to measure the review process in a more powerful way might be beneficial in getting good practices to stick.

You can see related commentary by Lance Fortnow (<http://bit.ly/1HNfPm7>), Bert Huang (<http://bit.ly/1DpGf6L>), and Yisong Yue (<http://bit.ly/1zNvoqb>).

Mark Guzdial
"Rising Enrollment
Might Capsize
Retention and
Diversity Efforts"

<http://bit.ly/1J3lst0>
 January 19, 2015

Computing educators have been working hard at figuring out how to make sure students succeed in computer science classes—with measurable success. The best paper award for the 2013 SIGCSE Symposium went to a paper on how a combination of pair programming, peer instruction, and curriculum change led to dramatic improvements in retention (<http://bit.ly/1EB9mle>). The chairs award for the 2013 ICER Conference went to a paper describing how Media Computation positively impacted retention in multiple institutions over a 10-year period (<http://bit.ly/1AkPH2x>). The best paper award at ITICSE 2014 was a meta-analysis of papers exploring approaches to lower failure rates in CS undergraduate classes (<http://bit.ly/1zNrvBH>).

How things have changed! Few CS departments in the U.S. are worried about

retention right now; instead, they are looking for ways to manage rising enrollment threatening to undermine efforts to increase diversity in CS education.

Enrollments in computer science are skyrocketing. Ed Lazowska and Eric Roberts sounded the alarm at the NCWIT summit last May, showing rising enrollments at several institutions (see <http://tern.ch/1zxUho2> and charts at right). Indiana University's undergraduate computing and informatics enrollment has tripled in the last seven years (<http://bit.ly/1EBaX0K>). At the Georgia Institute of Technology (Georgia Tech), our previous maximum number of undergraduates in computing was 1,539 set in 2001. As of fall 2014, we have 1,665 undergraduates.

What do we do? One thing we might do is hire more faculty, and some schools are doing that. There were over 250 job ads for CS faculty in a recent month. I do not know if there are enough CS Ph.D.'s looking for jobs to meet this kind of growth in demand for our courses.

Many schools are putting the brakes on enrollment. Georgia Tech is limiting transfers into the CS major and minor. The University of Massachusetts at Amherst is implementing caps. The University of California at Berkeley has a minimum GPA requirement to transfer into computer science.

We have been down this road before. In the 1980s when enrollment spiked, a variety of mechanisms were put into place to limit enrollment (<http://bit.ly/1KkZ9hB>). If there were too few seats available in our classes, we wanted to save those for the "best and brightest." From that perspective, a minimum GPA requirement made sense. From a diversity perspective, it did not.

Even today, white and Asian males are more likely to have access to Advanced Placement Computer Science and other pre-college computing opportunities. Who is going to do better in the intro course: the female student trying out programming for the first time, or the male student who has already programmed in Java? Our efforts to increase the diversity of computing education are likely to be undermined by efforts to manage rising enrollment. *Students who get shut out by such limits are most often in demographic groups underrepresented in computer science.*

When swamped by overwhelming numbers, retention is not your first

priority. When you start allocating seats, students with prior experience look most deserving of the opportunity. Google is trying to help; it started a pilot program to offer grants to schools with innovative ideas to manage booming enrollment *without* sacrificing diversity (<http://bit.ly/16b292F>).

One reason to put more computer science into schools is the need for computing workers (<http://bit.ly/1DxYbN3>). What happens if kids get engaged by activities like the Hour of Code, then cannot get into undergraduate CS classes? The linkage between computing in schools and a well-prepared workforce will be broken. It is ironic our efforts to diversify computing may be getting broken by *too many* kids being sold on the value of computing!

It is great students see the value of computing. Now, we have to figure out how to meet the demand—without sacrificing our efforts to increase diversity. ■

John Langford is a principal researcher at Microsoft Research New York. Mark Guzdial is a professor at the Georgia Institute of Technology.

© 2015 ACM 0001-0782/15/04 \$15.00