Is POPL Mathematics or Science?

Report of the Program Chair at the Nineteenth POPL

Andrew W. Appel
Department of Computer Science
Princeton University

January 1992

There has been some discussion of the direction that POPL is heading; perhaps the conference is becoming too theoretical. But we can easily measure the degree to which a conference is theoretical:

- When writing a joint paper, mathematicians order the names of the authors alphabetically.
- Scientists (experimental physicists, biologists, etc.) use a different criterion to order the names of the authors: The one with the grant goes first, or last (depending on the field); or the one that did the work goes first; or the student goes first, or last; but in any case alphabetical order is not used.

Now, these observations may be overgeneralizations. But from now on we will take them as *axioms*, so that they are not subject to argument.

By studying the order of authors' names on the joint papers in a given conference proceedings, we can discover whether the conference is largely theoretical (alphabetical order) or practical (non-alphabetical). Of course, it could be that when k scientists write a joint paper, the order of the authors will come out alphabetical anyway. This will occur with probability $1/k!$.

Suppose that a conference has a particular community of authors of which a fraction p are practitioners (scientists) and $1-p$ are theorists (mathematicians). Then we can say that the probability of a given k-author paper having alphabetical order is

$$(1-p) \; + \; p/k!$$

If there are n_2 2-author papers, n_3 3-author papers, etc., then the probability of having exactly a_2 alphabetical 2-author papers, *and* exactly a_3 alphabetical 3-author papers, etc. is

$$L(p) = \prod_k \left((1-p) + \frac{p}{k!}\right)^{a_k} \left(p - \frac{p}{k!}\right)^{n_k - a_k} \binom{n_k}{a_k}$$

Now, suppose we have a conference proceedings in hand; we can look through the table of contents to find n_k and a_k, for $k = 2, 3, \ldots$ (unfortunately the single-author papers tell us nothing). Then we can estimate the most likely value of p to produce that proceedings. The statisticians would have us do this by taking the log of the formula above and equating with 0 its partial derivitive with respect to p, but the hacker's way is just to do it numerically: try values of p between 0 and 1 and plot the curve.

Furthermore, we can integrate under the curve (again numerically) to find the error bars for p. Let

$$T = \int_0^1 L(p)dp \quad p_{\max} = \max_{p=0}^{1} L(p)$$

Then we would like to find p_{lo} and p_{hi} so that the chance that $p_{\text{lo}} < p < p_{\max}$ is about $T/3$, and the chance that $p_{\max} < p < p_{\text{hi}}$ is about $T/3$. We can find these by numerical integration.

Now all that is required is to take a big pile of conference proceedings, enter the data into the computer, and plot the results. The conferences are:

POPL: Principles of Programming Languages

PLDI: Programming Language Design and Implementation

LFP: Lisp and Functional Programming

FPCA: Functional Programming Languages and Computer Architecture

PPOPP: Principles and Practice of Parallel Programming (formerly PPEALS)

PEPM: Partial Evaluation and Semantics-Based Program Manipulation

OOPSLA: Object-Oriented Programming: Systems, Languages, and Applications

ASPLOS: Architectural Support for Programming Languages and Operating Systems

FOCS: Foundations of Computer Science

STOC: Symposium on the Theory of Computing

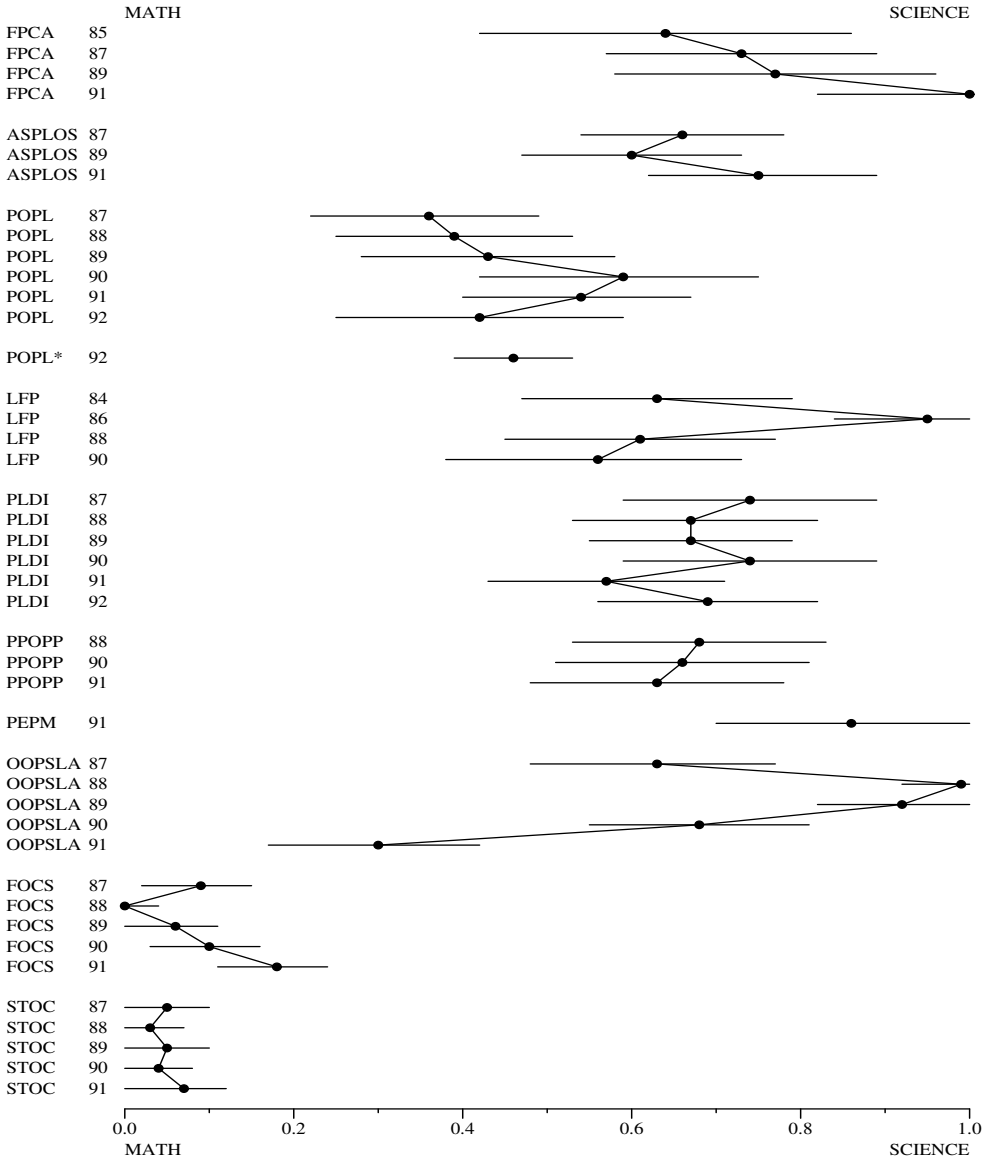

Figure 1: Math vs. Science in various conferences

The results (figure 1) out to be very informative. What we learn is that ASPLOS and PLDI are mostly practical (70%), and that FOCS and STOC are almost entirely theoretical (10% applied), although there is a statistically significant trend toward the practical in FOCS. POPL is about 40% applied, but has varied quite a bit from year to year.

None of these results are very surprising. The graph of OOPSLA is quite interesting: in 1988 it was 100% applied, and now it is more theoretical than POPL! I haven't been to an OOPSLA conference, so this was just a guess; but after presenting these results at the POPL conference, I was told that OOPSLA had indeed been much more theoretical this year. It's nice to see that my new measure of practicality has predictive value!

On the other hand, I'm not sure I'd agree that FPCA has been getting more and more practical. However, the error bars are quite large on that conference because of the large number of single-author papers.

The entry labelled POPL* covers the *submissions* to this year's conference: clearly, the accepted papers were about as practical as the submitted ones (the difference is not statistically significant, as one can tell by the error bars).

The goal of the research presented in this report was to provide some laughs at POPL '92; I am happy to say that in this respect the research was entirely successful.