

A somewhat open letter to Ben Kuipers

Dear Ben,

you speculated why theorists are tempted to regard experimental CS as a second-class effort. I think there is more to it. By virtue of the calling of the universities, academic disciplines are teachable disciplines and the primary purpose of academic research is to contribute to the improvement of teachable disciplines. A major objection to "experimental CS" is that it is a relatively very expensive and inefffficient way of improving our teachable discipline: the equipment and its maintenance are expensive, the programs to be written cost an inordinate amount of effort, and the contribution to what we can and should teach is meagre. [Lest you misunderstand me: I did not start as a "theorist" and still refuse to be characterized as such. For 15 years I was in the fortunate situation that I could try out novelties using each time the most powerful equipment then available in my country. But a quarter of a century ago, I saw funding problems lurking around the corner, and I bought my intellectual freedom by concentrating on work that would not require equipment or assisting staff. My now not using machines is a direct consequence of my decision to reduce the expense and increase the effectiveness of my research.]

The whole idea of experimental CS has one of its seeds in the position paper by Newell, Simon and Perlis, the tenor of which was "Facts breed science; computers exist, ergo!". The weakness of their argument is that one can substitute "postage stamps" for "computers" and then use it to defend a Department of Philately. Having listened to so-called "experimental computer scientists", I have come to the conclusion that the term "experimental CS" is a gross misnomer.

In the natural sciences, experiments are indispensable to check whether informal nature provides an acceptable model for your theory. With modern technology we can easily build artefacts that --though formally defined-- are of such a structure that a derivation of their relevant properties is way beyond our analytical abilities, but can such artefacts play the role of "informal nature", as mentioned in the previous sentence?

Hardly. You see, experiments involve measuring, but to think that the measurements constitute the experiment is like equating composing with the writing of the score. It is always the theory at stake that lends the experiment its significance. Think of the experiment of Michelson and Morley, refuting a "stationary aether" or the one of Millikan, which to all intents and purposes only admits models with discrete electrons. That, in passing, the speed of light and the charge of the electron were measured, was of secondary significance. (This "secondary" is no exaggeration: you can teach the whole of the Theory of Relativity, complete with all the experimental evidence, without ever mentioning the actual value of the speed of light: just c will do.) What I have seen so far in "experimental CS" is measuring without any theory that could be refuted or supported by the outcome of the measurements; consequently, the latter are at most of very local significance and hardly contribute to "the improvement of our teachable discipline". In all cases I remember, "performance measurement" would have been a much more appropriate term than "experimental CS".

The answer to the question to what extent we should try to teach our students how to measure performance probably depends on the extent to which we think that a first-class university should be involved in vocational training.

In the early 1960s, I have been involved in what now perhaps would be called "experimental CS". Friends were designing a floating point arithmetic unit and wanted to know the probability distribution of the difference of the exponents in the floating point representation of addenda. (They would like to know this in order to decide what additional hardware would be defensible for speeding up the mantissa alignment.) The brilliant physicist I then thought I still was tried to solve this problem theoretically, but no matter what ingenious ensemble of almost infinitely many computations I created, I got nowhere. This, of course, should have made me very suspicious, but it did not, and, unable to determine the distribution on theoretical grounds I decided to measure it. I had access to the ALGOL 60 implementation of the Mathematical Centre in Amsterdam and could mechanically observe all floating point additions performed in that institute. I collected my statistics. Due to storage constraints I could only collect the histograms per program, rather than have the machine produce the collective histogram for a full day's work. As a result I saw that between programs the histograms varied wildly! The whole idea of "an average program" was a complete fiction and the probability distribution sought just did not exist. And then, somewhat belatedly, I realized that I had expected a meaningful result from averaging operators, but in the absence of the equivalence of the Theorem of Liouville that lends significance to (at least classical) statistical mechanics. Without such an underlying theorem, averaging has about as much significance as a Gallup poll.

You mentioned "investigation of a general phenomenon by empirical study of particular instances of that phenomenon", and I think I can join you in that, but would like to point out that something "specific" can only be viewed as a particular instance of something more "general" after the introduction of the concepts in which the generalisation or particularization can be expressed and reasoned about. Otherwise the whole process is intellectually as shallow as trying to answer the question "How do we generalize broccoli?". Recently we have been exposed to lectures by all sorts of computing experimentalists. I don't remember if you attended all of them, but if you did you could have observed that in the vast majority of cases the speaker talked for more than an hour without introducing a single new concept; no wonder the talks were not fully satisfactory. In fact, the way in which some of the distinguished speakers disqualified themselves can not be expected to spread the opinion that experimental CS is a first-class effort.

* * *

The other weekend I spent a few hours reading in "Über Statistische Methoden in der Theorie der Quanta" (On statistical methods in quantum theory), the 1927 Ph.D. thesis of George Eugène Uhlenbeck, a thesis of which I happen to own a copy. I read it for a variety of reasons, among others as a reminder of the meticulous care with which our great scientists proceed. It was very refreshing.

Greetings!

Yours ever,

Edsger

Austin, Tuesday 19 November 1991

prof.dr.Edsger W.Dijkstra
 Department of Computer Sciences
 The University of Texas at Austin
 Austin, TX 78712 - 1188
 USA