



Science Has *Four* Legs

AS AN EDITOR of *The Fourth Paradigm* (<http://research.microsoft.com/en-us/collaboration/fourthparadigm/default.aspx>, Microsoft Research, Redmond, WA, 2009) and someone who subscribes to Jim Gray's vision that there are now four fundamental scientific methodologies, I feel I must respond to Moshe Y. Vardi's Editor's Letter "Science Has Only Two Legs" (Sept. 2010).

First, I should explain my qualifications for defending the science-has-four-legs premise. From 1964, beginning as a physics undergraduate at Oxford, until 1984, when I moved from physics to the Electronics and Computer Science Department, I was a working natural scientist. My Ph.D. is in theoretical particle physics, and, in my research career, I worked extensively with experimentalists and spent two years at the CERN accelerator laboratory in Geneva. In computer science, my research takes in all aspects of parallel computing—architectures, languages, and tools, as well as methodologies for parallelizing scientific applications—and more recently the multi-core challenge. From 2001 to 2005, before I joined Microsoft, I was Director of the U.K.'s eScience Core Program, working closely with scientists of all descriptions, from astronomers and biologists to chemists and environmental scientists. Here at Microsoft Research, I still work with practicing scientists.

I therefore have some relevant experience on which to ground my argument. By contrast, though Vardi has had a distinguished career in mathematics and computer science (and has done a great job with *Communications*), he has not, as far as I know, had much direct involvement with the natural sciences.

It is quite clear that the two new scientific paradigms—computational and data-intensive—do not displace experiment and theory, which remain as relevant as ever. However, over the

past 50 years it is equally clear that computational science has emerged as a third methodology with which we now explore problems that are simply inaccessible to experiment. To do so, scientists need (along with their knowledge of experiment and theory) training in numerical methods, computer architecture, and parallel programming. It was for this reason that Physics Nobel Prize laureate Ken Wilson in 1987 called computational science the "third paradigm" for scientific discovery. He was investigating quantum chromo-dynamics, or QCD, describing the fundamental equations between quark and gluon fields behind the strong nuclear force. No analytic solution is possible for solving these equations, and the only option is to approximate the theory on a space-time lattice. Wilson pioneered this technique, using supercomputers to explore the predictions of QCD in the physical limit when the lattice spacing tends to zero. Other examples of such computational exploration, including galaxy formation and climate modeling, are not testable through experiment in the usual sense of the word.

In order to explore new techniques for storing, managing, manipulating, mining, and visualizing large data sets, Gray felt the explosive growth of scientific data posed profound challenges for computer science (see Fran Berman's "Got Data? A Guide to Data Preservation in the Information Age," *Communications*, Dec. 2008). He therefore spent much of his last years working with scientists who were, as he said, "drowning in data." Working with astronomers allowed him the luxury of experimentation, since, he said, their data had "absolutely no commercial value." Similarly, the genomic revolution is upon us, and biologists need powerful tools to help disentangle the effects of multiple genes on disease, develop new vaccines, and design effective drugs. In environmental science, data from large-scale sensor networks is begin-

ning to complement satellite data, and we are seeing the emergence of a new field of environmental informatics. Scientists require significant help not only in managing and processing the raw data but also in designing better workflow tools that automatically capture the provenance involved in producing the data sets scientists actually work with.

On the other hand, "computational thinking" attempts to demonstrate the power of computer science ideas not just in science but also in many other aspects of daily life. However, despite its importance, this goal should not be confused with the emergence of the two new methodologies scientists now need to assist them in their understanding of nature.

Tony Hey, Seattle

Author's Response:

Hey and I are in violent agreement that science today is thoroughly computational. What I fail to see is why this requires it to sprout new legs. In fact, theory in science was mathematical way before it was computational. Does that make mathematics another leg of science?

Experimental science always relied on statistical analysis. Does that make statistics another leg of science? Science today relies on highly complex theoretical models, requiring analysis via computation, and experimental setups that yield massive amounts of data, also requiring analysis via computation. So science is thoroughly computational but still has only two legs—theory and experiment.

Moshe Y. Vardi, Editor-in-Chief

Let Patients Participate in Their Own Care

In his article "Computers in Patient Care: The Promise and the Challenge" (Sept. 2010), Stephen V. Cantrill, M.D., offered seven compelling arguments for integrating health information technology (HIT) into

clinical practice. However, he missed one that may ultimately surpass all others—making medical data meaningful (and available) to patients—so they can be more informed partners in their own care.

Dr. Cantrill was not alone in appreciating the value of HIT this way. Patient-facing electronic data presentation is consistently overlooked in academic, medical, industrial, and political discussions, likely because it's much more difficult to associate financial value with patient engagement than with measurable inefficiencies in medical practice.

Perhaps, too, computer scientists have not let patients take advantage of the growing volume of their own electronic medical data; allowing them only to, say, download and print their medical histories is important but insufficient. Medical data is (and probably should be) authored by and for practitioners, and is thus beyond the health literacy of most patients. But making medical data intuitive to patients—a problem that's part pedagogy, part translation, part infrastructure, and part design—requires a collaborative effort among researchers in human-computer interaction, natural language processing, visualization, databases, and security. The effort also represents a major opportunity for CS in terms of societal impact. Its omission is indicative of just how much remains to be done.

Dan Morris, Redmond, WA

Release the Code

About the software of science, Dennis McCafferty's news story (Oct. 2010) asked "Should Code Be Released?" In the case of climate science code, the Climate Code Foundation (<http://climatecode.org/>) answers with an emphatic yes. Rebuilding public trust in climate science and support for policy decisions require changes in the transparency and communication of the science. The Foundation works with climate scientists to encourage publication of all climate-science software.

In a *Nature* opinion piece "Publish Your Computer Code: It Is Good Enough" (Oct. 13, 2010, <http://www.nature.com/news/2010/101013/>

full/467753a.html), I argued that there are powerful reasons to publish source code across all fields of science, and that software is an essential aspect of the scientific method despite failing to benefit from the system of competitive review that has driven science forward for the past 300 years. In the same way software is part of the scientific method, source code should be published as part of the method description.

As a reason for not publishing software, McCafferty quoted Alan T. DeKok, a former physicist, now CTO of Mancala Networks, saying it might be "blatantly wrong." Surely this is a reason, perhaps the main one, that software *should* be published—to expose errors. Science progresses by testing new ideas and rejecting those that are wrong.

I'd also like to point out a glaring red herring in McCafferty's story—the suggestion that a policy in this area could undermine a modern-day Manhattan Project. All design and method descriptions from that project were top secret for years, many to this day. Such secrecy would naturally apply to any science software of similar military importance.

Nick Barnes, Staines, U.K.

The Brain's Inner Computer Is Analog

I continue to be amazed by the simplistic approach pursued by computer scientists trying to understand how the brain functions. David Lindley's news article "Brains and Bytes" (Sept. 2010) came tantalizingly close to an epiphany but didn't quite express what to me is fundamentally wrong with most research in the field. There is an appreciation of the statistical nature of the brain's functioning at the microscopic, cellular level, a realization that complete predictability is not only not achievable but actually completely inappropriate.

Lindley referred to an event ("neural firing") as a binary process, despite being statistical in its occurrence. Lacking personal experience (so unfettered by knowledge), I claim this represents the fundamental obstacle to achieving a true understanding of how the brain works. A neuron firing

or a synapse transmitting the result is neither binary nor random; rather, the shape and strength of the "signal" are critical in achieving understanding, and are, for the most part, ignored.

Many researchers seem precommitted to a view defined by digital processes coupled with statistical unpredictability. Time to return to the Dark Ages of computing when the brain's cellular components were not statistically imperfect digital devices. They were and are analog, a word Lindley left out of the article, even though some of his descriptions of the cellular functions cried out for such a characterization.

R. Gary Marquart, Austin, TX

Communications welcomes your opinion. To submit a Letter to the Editor, please limit your comments to 500 words or less and send to letters@cacm.acm.org.

© 2010 ACM 0001-0782/10/1200 \$10.00

Coming Next Month in COMMUNICATIONS

*A Firm Foundation
for Private Data Analysis*

Using Simple Abstraction

*An Interview with
Fran Allen*

*Cloud Computing
at our Doorstep*

*Follow the Intellectual
Property*

Q&A with Ed Lazowska

*Technology Strategy
and Management*

The Business of Software

Law and Technology

ACM's FY10 Annual Report

And the latest news on nonlinear logic, haptic user interfaces, and AAI for developing nations.