opinion

The end of theory in science?

Philosopher Ludwig Wittgenstein (1889–1951) famously said that, "whereof one cannot speak, thereof one must be silent." Although his counsel was given in a specific context, it is actually good general advice that Chris Anderson, Editor-in-chief of *Wired* magazine, would do well to heed. On June 23, 2008, Anderson posted an article on *Wired's* website, "The end of theory: the data deluge makes the scientific method obsolete," from which it is perfectly clear that he doesn't understand much about either science or the scientific method.

Anderson's main point is that the modern era of 'petabyte' information and 'cloud' computing on the web is bypassing the 'hypothesize, model, test' procedure of science because scientific theorizing simply cannot cope with the deluge of data. Here is an excerpt: "Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity [...] the numbers speak for themselves." Actually, the point may or may not be why people do things as opposed to what they do—it just depends on one's interest. But the numbers, contrary to Anderson's bold assertion, do not, in fact, speak for themselves. As Charles Darwin (1809–1882) famously put it: "How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service!"

If Anderson had been talking about advertizing and how companies should select their targets or fine-tune their merchandizing, he would have been right. But he makes it clear that, "[t]he big target here isn't advertising, though. It's science. [...] Scientists are trained to recognize that correlation is not causation [...] There is now a better way. Petabytes allow us to say: 'Correlation is enough.' We can stop looking for models."

But, if we stop looking for models and hypotheses, are we still really doing science? Science, unlike advertizing, is not about finding patterns—although that is certainly part of the process—it is about finding explanations for those patterns. In fact, it is easy to argue that Anderson is wrong even about advertizing. While advertizers might not be interested in theories of human behaviour—actually, they are, and they use them to the best of their abilities—they still collect and organize data in a particular way, which is what one does when using Google's petabyte-sized cloud, and this must involve the formulation and testing of hypotheses. Why collect certain pieces of information rather than others? Why use certain keywords to organize the search rather than others? Every choice we make in that respect is a reflection of an, often unstated, set of assumptions and hypotheses about what we want and expect from the data. Without models, mathematical or conceptual, data are just noise.

Let's take Anderson's example of what is wrong with science: theoretical physics. In the article, he writes that: "The reason physics has drifted into theoretical speculation about *n*-dimensional grand unified models over the past few decades [...] is that we don't know how to run the experiments that would falsify the hypotheses—the energies are too high, the accelerators too expensive, and so on." While this is true, the problem here is one of insufficient information—what philosophers call the underdetermination of theories by the data—not of too much information, which would be Anderson's contention for why science is in trouble.

Anderson goes on to propose a positive example of the new science he envisions: molecular biology done *a la* Craig Venter, the entrepreneur scientist. According to Anderson, "Venter has advanced biology more than anyone else of his generation," and has done so, among other things, by conducting high throughput searches of genomes in the ocean. In fact, Venter has simply collected buckets of water, filtered the material and put the organic content through his high-speed genomic sequencing machines. The results are interesting, including the discovery that there are thousands of previously unknown bacterial species. But, as Anderson points out, "Venter can tell you almost nothing about the species he found. He doesn't know what they look like, how they live, or much of anything else about their morphology. He doesn't even have their entire genome. All he has is a statistical blip—a unique sequence that, being unlike any other sequence in the database, must represent a new species." Which means that Venter has succeeded in generating a large amount of data—in response to a specific question, by the way: how many distinct, species-level genome sequences can be found in the oceans? This will surely provide plenty of food for thought for scientists, and a variety of ways to test interesting hypotheses about the structure of the biosphere, the diversity of bacterial life, and so on. But, without those hypotheses to be tested, Venter's data are going to be a useless curiosity, far from being the most important contribution to science in this generation.

Anderson boldly closes his piece of epistemic bravado by stating that: "The new availability of huge amounts of data [...] offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all." Yet, science advances only if it can provide explanations, failing which, it becomes an activity more akin to stamp collecting. Now, there is an area where petabytes of information can be used for their own sake. But please don't call it science.

Massimo Pigliucci is a professor of philosophy at the City University of New York. His research focuses on the structure of evolutionary theory (www.platofootnote.org). His most recent book (with Jonathan Kaplan) is **Making Sense of Evolution.**
E-mail: massimo@platofootnote.org