By: John C. Mather, www.edge.org essay

You may not like it! But artificial intelligence jumped a bit closer this year with the development of "Bayesian Program Learning," by Lake, Salakhutdinov, and Tenenbaum, published in *Science*. It's news because for decades I've been hearing about how hard it is to achieve artificial intelligence, and the most successful methods have used serious brute force. Methods based on understanding the symbols and logic of things and language have had a tough time. The challenge is to invent a computer representation of complex information, and then enable a machine to learn that information from examples and evidence.

Lake et al. give a mathematical framework, an algorithm, and a computer code that implements it, and their software has learned to read 1623 handwritten characters in 50 languages as well as a human being. They say "Concepts are represented as simple probabilistic programs—that is, probabilistic generative models expressed as structured procedures in an abstract description language." Also, a concept can be built up by re-using parts of other concepts or programs. The probabilistic approach handles the imprecision of both definitions and examples. (Bayes' theorem tells us how to compute the probability of something complicated, if we know the probabilities of various smaller things that go into the complicated thing.) Their system can learn very quickly, sometimes in one shot, or from a few examples, in a human-like way, and with human-like accuracy. This ability is in dramatic contrast to competing methods depending on immense data sets and simulated neural networks, which are always in the news.

So now there are many new questions: How general is this approach? How much structure do humans have to give it, to get it started? Could it really be superior in the end? Is this how living intelligent systems work? How could we tell? Can this computer system grow enough to represent complex concepts that are important to humans in daily life? Where are the first practical applications?

This is a long-term project, without any obvious limits to how far it can go. Could this method be so efficient that it doesn't take a super-duper supercomputer to achieve or at least represent artificial intelligence? Insects do very well with a tiny brain after all. More generally, when do we get really accurate transcriptions of multi-person conversations, instantaneous machine language translation, scene recognition, face recognition, self-driving cars, self-directed UAVs safely delivering packages, machine understanding of physics and engineering, machine representation of biological concepts, and machine ability to read the Library of Congress and discuss it in a philosophy or history class? When will my digital assistant really understand what I want to do, or tell me what I ought to do? Is this how the Intelligent Mars Rover will hunt for signs of life on Mars? How about military offense and defense? How could this system implement Asimov's three laws of robotics, to protect humans from robots? How would you know if you should trust your robot? When will people be obsolete?

I'm sure many people are already working on all of these questions. I see many opportunities for mischief, but the defense against the dark arts will push very rapid progress too. I am both thrilled and frightened.