The Believers

The hidden story behind the code that runs our lives

By Paul Voosen February 23, 2015 Premium

Magic has entered our world. In the pockets of many Americans today are thin black slabs that, somehow, understand and anticipate our desires. Linked to the digital cloud and satellites beyond, churning through personal data, these machines listen and assist, decoding our language, viewing and labeling reality with their cameras. This summer, as I walked to an appointment at the University of Toronto, stepping out of my downtown hotel into brisk hints of fall, my phone already had directions at hand. I asked where to find coffee on the way. It told me. What did the machine know? How did it learn? A gap broader than any we've known has opened between our use of technology and our understanding of it. How did the machine work? As I would discover, no one could say for certain. But as I walked with my coffee, I was on the way to meet the man most qualified to bridge the gap between what the machine knows and what you know.

Geoffrey Hinton is a torchbearer, an academic computer scientist who has spent his career, along with a small band of fellow travelers, devoted to an idea of artificial intelligence that has been discarded multiple times over. A brilliant but peripheral figure. A believer. A brusque coder who had to hide his ideas in obscure language to get them past peer review. A devotee of the notion that, despite how little we understand the brain, even a toy model of it could present more computational power and flexibility than the rigid logic or programmed knowledge of traditional artificial intelligence. A man whose ideas and algorithms might now help power nearly every aspect of our lives. A guru of the artificial neural network.

Such networks, which have been rebranded "deep learning," have had an

unparalleled ascent over the past few years. They've [hit the front page](#) of *The New York Times*. Adept at processing speech, vision, and other aspects of the messy interface with humanity that has been sped up by ubiquitous mobile devices, nets have been embraced by Google, Facebook, Microsoft, Baidu, and nearly any other tech leader you can imagine. At these companies, neural nets have proved an efficient way to soak up vast amounts of data and make highly valuable predictions from it: How do you make a data center more energy efficient? Will this user want to buy a car soon? Tech companies compete fiercely for every coder who shows an aptitude in developing neural nets, often luring them away from careers in academe. Last year, for Google, that included reportedly spending more than $400-million on a company, DeepMind, with no products, only [a way of integrating memory](#) into learning algorithms. And before that, Google bought Hinton's services for an undisclosed sum.

There's seemingly no crevice of technology that hasn't felt the creep of deep learning. Over the months, announcements pile up in my inbox: Deep learning that identifies [autism-risk genes](#). Deep learning that writes automated captions for [pictures](#) and [video.](#) Deep learning to [identify particles](#) in the Large Hadron Collider. Deep learning to [guide our cars](#) and robots.

"If you want to understand how the mind works, ignoring the brain is probably a bad idea."

With each announcement, deep learning has nudged the notion of artificial intelligence back into the public sphere, though not always to productive ends. Should we worry about the robot revolution to come? Spoiler alert: not right now; maybe in 50 years. Are these programmers foolish enough to think they're actually mimicking the brain? No. Are we on the way to truly intelligent machines? It depends on how you define intelligence. Can deep learning live up to its hype? Well ...

Such a clamor has risen around deep learning that many researchers warn that if they don't deliver on its potential, they risk a backlash against all of artificial intelligence. "It's damaging," says Yann LeCun, a professor at New York University who now directs [Facebook's AI research.](#) "The field of AI died three or four times because of hype."

Several of these deaths came at the hands of artificial neural networks. In the 1960s and again in the 1980s, neural nets rose like a rocket, only to fall to earth once the limitless dreams of their creators met the limits of transistors. During those dark days, the few devoted researchers, like Hinton and LeCun, were down in the "rat holes," ignored by the academic world, one longtime Hinton collaborator told me. Few would have expected a third ascent. Many still fear another crash.

Hinton, however, is all confidence. He had invited me to Toronto to learn about this new era's deep history. For a decade, he's run a weeklong [summer school](#) on neural nets; I stopped by while it was under way. It was a hot day of dry presentations and young men, mostly men, with overflowing hopes packed into overflowing rooms. I found Hinton in his office, which he's kept despite becoming emeritus. A bad back leaves him standing; when he travels to Google's headquarters in California for half the year, he goes by train. Decorating his door were handwritten digits, indecipherable, pulled from [a data set](#) that provided some of neural networks' earliest successes.

It's hard for Hinton, 67, not to feel a bit pleased with himself. After a lifetime on the periphery, he now has a way of connecting with nearly anyone he meets. For example, when in Toronto, he works out of Google's office downtown, which is filled with advertising employees. He's the only researcher. Occasionally, a curious employee sidles up and asks, "What do you do?"

"Do you have an Android phone?" Hinton replies.

"Yes."

"The speech recognition is pretty good, isn't it?"

"Yes."

"Well, I design the neural networks that recognize what you say."

The questioner nearly always pauses in thought.

"Wait, what do you mean?"

For nearly as long as we've attempted to create "thinking" computers, researchers have argued about the way they should run. Should they imitate how we imagine the mind to work, as a Cartesian wonderland of logic and abstract thought that could be coded into a programming language? Or should they instead imitate a drastically simplified version of the actual, physical brain, with its web of neurons and axon tails, in the hopes that these networks will enable higher levels of calculation? It's a dispute that has shaped artificial intelligence for decades.

One pioneer of brain imitation, in the late 1950s, was Frank Rosenblatt, a psychologist at the Cornell Aeronautical Laboratory. He was inspired by the work of Donald O. Hebb, who a decade earlier had predicted how learning might work: As one neuron fires and activates another, repeatedly, the cells improve their joint efficiency. "The cells that fire together, wire together," as cognitive scientists like to say. This simple idea, Rosenblatt thought, was enough to build a machine that could learn to identify objects.

Build it he did: You can see parts of the Perceptron, as he called it, in the Smithsonian. Its operation was simple. Taking up an entire lab room, it worked in three layers. At one end, a square cluster of 400 light sensors simulated a retina; the sensors connected multiple times to an array of 512 electrical triggers, each of which fired, like a neuron, when it passed a certain adjustable threshold of excitement. These triggers then connected to the last layer, which would signal if an object matched what the Perceptron

had been trained to see.

Trained is the operative word: The Perceptron was not programmed, but trained. It could not learn on its own. Rosenblatt created a formula that calculated how much the Perceptron was wrong or right, and that error could then be traced back and individually changed in those 512 triggers. Tweak these weights enough, and the Perceptron could begin to recognize very basic patterns, such as standardized letter shapes.

It was a thrilling development, and Rosenblatt wasn't afraid to share it. In the summer of 1958, he held a news conference with his sponsor, the U.S. Navy. As so often happens in science, he began to talk about the future. To researchers then, he sounded foolish; heard today, prescient. *The New York Times* caught [the gist of it](#):

The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence. … Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech and writing in another language, it was predicted.

Rosenblatt's fame irked his peers, many of whom had opted to pursue rules-based artificial intelligence; both sides were chasing after the same military research dollars. Most prominently, Marvin Minsky and Seymour Papert, two eminent computer scientists at the Massachusetts Institute of Technology, sought to replicate the Perceptron and expose its flaws, [culminating](#) in a 1969 book seen as causing a near-death experience for neural nets. The Perceptron had intrinsic limitations, they said. Most basically, it could not learn "exclusive or," a basic bit of logic that holds true only if one input is true and the other false.

Learning that function would require an additional layer in the Perceptron. But no one could figure out a biologically plausible way to calculate and transmit the adjustments for such a "hidden" layer. The neural net

compressed information, in effect, which could not then be retrieved. It felt like time—it ran only forward. The learning had stopped, and the grant money vanished. Minsky and Papert had won.

Frustrated, Rosenblatt found other outlets. He became fascinated by a project attempting to show that brain cells transplanted from one rat to another would retain memory. The work didn't last long—he died young, in a 1971 sailing accident, alone, on his birthday. It seemed the neural network would die with him.

No one told Geoff Hinton, who after bouncing around in college among chemistry, physics, physiology, philosophy, and psychology, managed to enroll, in 1972, in a graduate program in artificial intelligence at the University of Edinburgh.

Hinton is the son of a peripatetic British clan whose members tended to do what they thought best. One great-great-grandfather was [George Boole](), whose algebra became the basis of the computer age, including that exclusive-or that defied Rosenblatt; another owned a Victorian sex club. His grandfather ran a mine in Mexico, and his father was an entomologist: "He thought things with six legs were much more interesting than things with two legs."

As a teenager, Hinton became fascinated with computers and brains. He could build electrical relays out of razor blades, six-inch nails, and copper wire in 10 minutes; give him an hour, and he'd give you an oscillator.

His view then was the same he has today: "If you want to understand how the mind works, ignoring the brain is probably a bad idea." Using a computer to build simple models to see if they worked—that seemed the obvious method. "And that's what I've been doing ever since."

This was not an obvious view. He was the only person pursuing neural nets in his department at Edinburgh. It was hard going. "You seem to be

intelligent," people told him. "Why are you doing this stuff?"

Hinton had to work in secret. His thesis couldn't focus on learning in neural nets; it had to be on whether a computer could infer parts, like a human leg, in a picture. His first paper on neural nets wouldn't pass peer review if it mentioned "neural nets"; it had to talk about "optimal networks." After he graduated, he couldn't find full-time academic work. But slowly, starting with a 1979 conference he organized, he found his people.

"We both had this belief," says Terrence J. Sejnowski, a computational neurobiologist at the Salk Institute for Biological Studies and longtime Hinton collaborator. "It was a blind belief. We couldn't prove anything, mathematical or otherwise." But as they saw rules-based AI struggle with things like vision, they knew they had an ace up their sleeve, Sejnowski adds. "The only working system that could solve these problems was the brain."

Hinton has always bucked authority, so it might not be surprising that, in the early 1980s, he found a home as a postdoc in California, under the guidance of two psychologists, David E. Rumelhart and James L. McClelland, at the University of California at San Diego. "In California," Hinton says, "they had the view that there could be more than one idea that was interesting." Hinton, in turn, gave them a uniquely computational mind. "We thought Geoff was remarkably insightful," McClelland says. "He would say things that would open vast new worlds."

They held weekly meetings in a snug conference room, coffee percolating at the back, to find a way of training their error correction back through multiple layers. Francis Crick, who co-discovered DNA's structure, heard about their work and insisted on attending, his tall frame dominating the room even as he sat on a low-slung couch. "I thought of him like the fish in *The Cat in the Hat*," McClelland says, lecturing them about whether their ideas were biologically plausible.

The group was too hung up on biology, Hinton said. So what if neurons

couldn't send signals backward? They couldn't slavishly recreate the brain. This was a math problem, he said, what's known as getting the gradient of a loss function. They realized that their neurons couldn't be on-off switches. If you picture the calculus of the network like a desert landscape, their neurons were like drops off a sheer cliff; traffic went only one way. If they treated them like a more gentle mesa—a sigmoidal function—then the neurons would still mostly act as a threshold, but information could climb back up.

While this went on, Hinton had to leave San Diego. The computer-science department had decided not to offer him a position. He went back to Britain for a lackluster job. Then one night he was startled awake by a phone call from a man named Charlie Smith.

"You don't know me, but I know you," Smith told him. "I work for the System Development Corporation. We want to fund long-range speculative research. We're particularly interested in research that either won't work or, if it does work, won't work for a long time. And I've been reading some of your papers."

Hinton won $350,000 from this mysterious group. He later learned its origins: It was a subsidiary of the nonprofit RAND Corporation that had ended up making millions in profit by writing software for nuclear missile strikes. The government caught them, and said they could either pay up or give the money away—fast. The grant made Hinton a much more palatable hire in academe.

Back in San Diego, Rumelhart kept on the math of their algorithm, which they started calling back-propagation. When it was done, he simulated the same exclusive-or that had defied Rosenblatt. He let it run overnight. When he returned the next morning, the neural net had learned.

By the late 1980s, neural nets were everywhere. They were back in *The New York Times,* which reviewed a technical book written by the San Diego team.

Companies promised to solve a fleet of problems with neural nets. Even Hollywood took notice: "My CPU is a neural net processor," Arnold Schwarzenegger's robotic terminator said. "A learning computer."
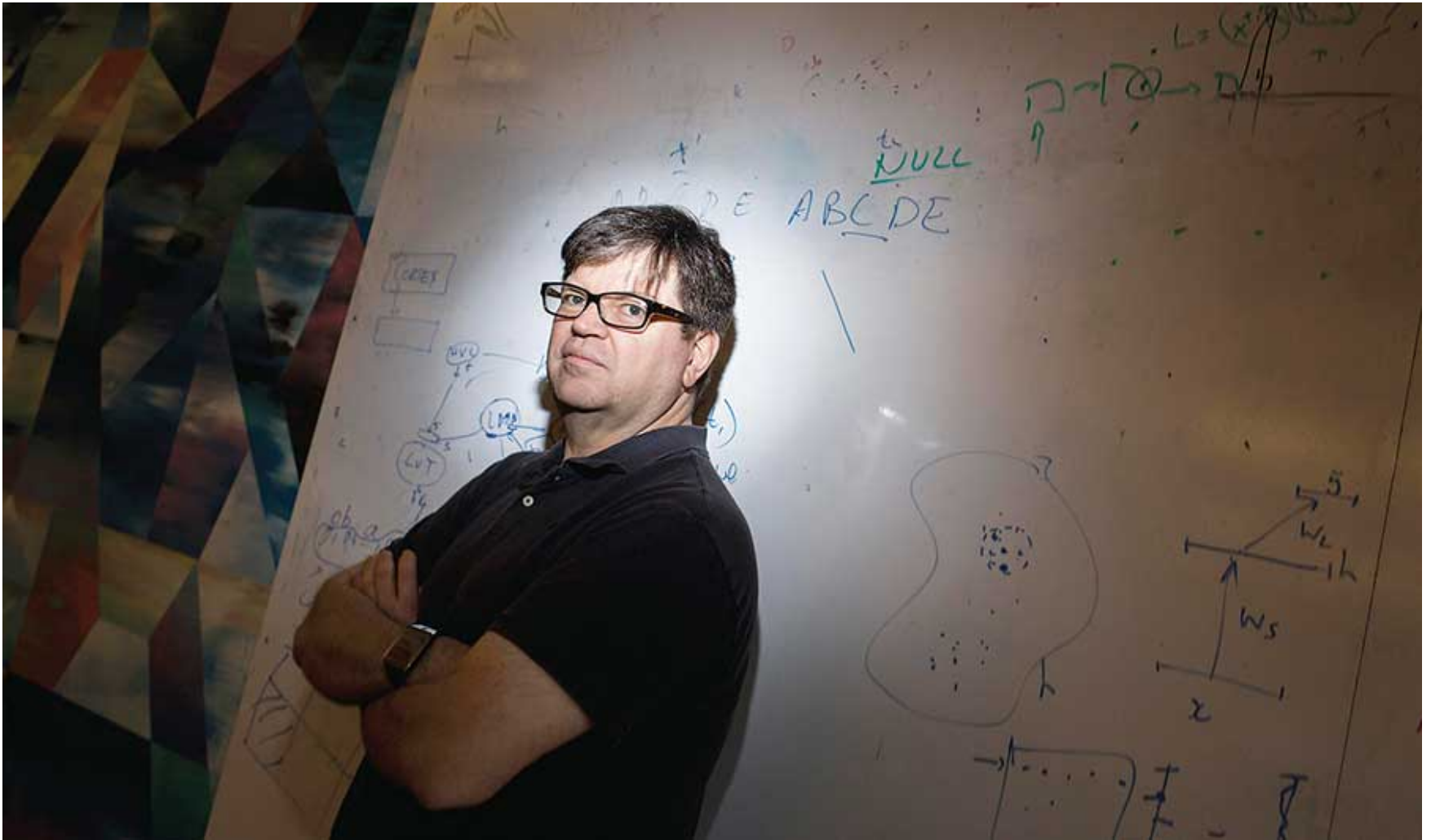
Mark Abramson for The Chronicle

Yann LeCun, a professor at NYU, also leads Facebook's artificial-intelligence laboratory.

Hinton spent a few years at Carnegie Mellon University. With Rumelhart and Ronald J. Williams, he had shown that neural nets could learn multiple layers of features, essential to proving that complex calculations could arise from such networks. But he was dissatisfied with back-propagation, which, it turned out, several others, including LeCun, had also invented—it just didn't seem powerful enough. With Sejnowski, he developed a neural net modeled after the Boltzmann distribution, a bit of statistical physics that describes the probabilities for how matter shifts energy states under changing temperatures. (Think water turning to ice.) It was classic Hinton: He builds code from physical analogies, not pure math. It was a fertile time. Sejnowski remembers sitting in his kitchen, getting a call from Hinton: "Terry, I've figured out how the brain works," Hinton said. Over the last 30 years,

Sejnowski adds, Hinton has called and told him that a dozen times.

The world didn't join Hinton in that excitement for long. The research hit new walls. Neural nets could learn, but not well. They slurped up computing power and needed a bevy of examples to learn. If a neural net failed, the reasons were opaque—like our own brain. If two people applied the same algorithm, they'd get different results. Engineers hated this fickleness, says Facebook's LeCun. This is too complicated, they said, therefore the people who use it must believe in magic. Instead, coders opted for learning algorithms that [behaved predictably](#) and seemed to do as well as back-propagation.

As they watched neural nets fade, they also had to watch Rumelhart, the man most responsible for their second wave, decline. He was slowly succumbing to Pick's disease, a rare dementia that, McClelland suggests, may arise from overusing the neurons in the brain. (He died in 2011.) The Cognitive Science Society began offering an award in Rumelhart's honor in 2001; Hinton was its first recipient.

The field lost its vision, says Yoshua Bengio, a professor at the University of Montreal who, in the 1990s, joined Hinton and LeCun as a neural-net partisan. Though a neural net LeCun had modeled after the visual cortex was reading up to 20 percent of all U.S. bank checks, no one talked about artificial intelligence anymore. "It was difficult to publish anything that had to do with neural nets at the major machine-learning conferences," Bengio told me. "In about 10 years, neural nets went from the thing to oblivion."

A decade ago, Hinton, LeCun, and Bengio conspired to bring them back. Neural nets had a particular advantage compared with their peers: While they could be trained to recognize new objects—supervised learning, as it's called—they should also be able to identify patterns on their own, much like a child, if left alone, would figure out the difference between a sphere and a cube before its parent says, "This is a cube." If they could get unsupervised learning to work, the researchers thought, everyone would come back. By

2006, Hinton had a paper out on ["deep belief networks,"](#) which could run many layers deep and learn rudimentary features on their own, improved by training only near the end. They started calling these artificial neural networks by a new name: "deep learning." The rebrand was on.

Before they won over the world, however, the world came back to them. That same year, a different type of computer chip, the graphics processing unit, became more powerful, and Hinton's students found it to be perfect for the punishing demands of deep learning. Neural nets got 30 times faster overnight. Google and Facebook began to pile up hoards of data about their users, and it became easier to run programs across a huge web of computers. One of Hinton's students interned at Google and [imported](#) Hinton's [speech recognition](#) into its system. It was an instant success, outperforming voice-recognition algorithms that had been tweaked for decades. Google began moving all its Android phones over to Hinton's software.

It was a stunning result. These neural nets were little different from what existed in the 1980s. This was simple supervised learning. It didn't even require Hinton's 2006 breakthrough. It just turned out that no other algorithm scaled up like these nets. "Retrospectively, it was a just a question of the amount of data and the amount of computations," Hinton says.

Hinton now spends half his year at Google's campus, preventing its engineers from traveling down dead ends from decades past. He's also exploring neural nets that might have been discarded as unworkable, and pursuing what he calls ["dark knowledge."](#) He often spends the full day coding, something he would never have been able to do as a professor. When I asked about the most productive part of his career, he replied without hesitation: "The next five years."

Google uses deep learning in dozens of products. When I visited Hinton this summer, it had just begun applying deep learning to language translation. Google has encoder and decoder networks for each language, which

convert each word into a big matrix of numbers that capture much of its meaning—the numbers for "cat" and "dog," say, will be much more similar than those for "dog" and "auburn." The English encoder passes those numbers to the French decoder, for example, which makes an overall prediction with those numbers, and then compares that prediction with word-by-word analysis as it goes, all the while comparing the results with known translations and back-propagating the errors. After a few months, it was already working well, Hinton said.

There's some irony that Hinton, such an iconoclast, is now working for a large company. But it's unavoidable. These companies have the tools to make deep learning work, and universities do not. During a coffee break at Hinton's summer school, I overheard a young academic griping about not getting training data from one company. After a few minutes, he added, "I'm going to Microsoft, so this won't be a problem for me soon."

"There is the slight danger that if enough big companies hire enough of the researchers, there won't be enough left in the universities to produce new students and to do the pure basic research," Hinton says. But the tech companies are aware of this problem, he adds. Google, for example, is eager for Bengio to keep on his basic research, Hinton says.

At Facebook, LeCun has been recreating a new version of Bell Labs, where he worked in the 1990s. They will publish their work, he promised, if at somewhat of a time delay. "I don't think academic research is going to be put out of existence," he adds. The tech rush for talent is creating more students than they're losing to industry. While he's wary of hype, he's also confident that deep learning is just getting started. "I wouldn't have been doing this for 20 years, against the better judgment of everybody, unless I

"We could have moved a lot faster, if it weren't for the ways of science as a human enterprise."

believed in these methods."

Bengio, for one, can't help but think back to all the grants not financed, the peer-review attacks from scientists invested in older approaches to computer vision or speech recognition. "We could have moved a lot faster, if it weren't for the ways of science as a human enterprise," he says. Diversity should trump personal biases, he says, "but humans tend to discard things they don't understand or believe in."

Now the neural-net researchers are dominant. Even faculty members at MIT, long a bastion of traditional artificial intelligence, are on board.

"We were little furry mammals scrambling under the feet of dinosaurs," Salk's Sejnowski says. "Basically, the little afraid mammals won. The dinosaurs are gone. It's a new era."

Many of the dreams Rosenblatt shared in his news conference have come true. Others, like computer consciousness, remain distant. The largest neural nets today have about a billion connections, 1,000 times the size of a few years ago. But that's tiny compared with the brain. A billion connections is a cubic millimeter of tissue; in a brain scan, it'd be less than a voxel. We're far from human intelligence. Hinton remains intrigued and inspired by the brain, but he knows he's not recreating it. It's not even close.

Speculation remains on what neural nets will achieve as they grow larger. Many researchers resist the notion that reasoning could ever evolve out of them. Gary F. Marcus, an NYU psychologist, [critiqued](#) the gains of deep learning [in several](#) *New Yorker* essays, to the point that Hinton pushed him to state what a neural net would have to do to impress him. [His answer?](#) Read this: "The city councilmen refused to give the demonstrators a license because they feared violence." *Who feared the violence?* If a neural net could answer that question, then they'd be on to something.

There's a deep irony in all this, Sejnowski adds. Deep learning is now one of

the most promising tools for exploiting the enormous databases stemming from neuroscience. "We started this thing to understand how the brain works," he says. "And it turns out the very tools we created, many not very brainlike, are the optimal tools to understand what neuroscience is doing."

It was a long day in Toronto. At one point during my visit, I noticed that Hinton had a program running on his laptop. Every few seconds, two black-and-white handwritten numbers flashed on screen, randomly overlaid. He was testing a new algorithm, seeing how well it did in detecting the two numbers despite the visual clutter.

Two new numbers appeared. His eyes turned mischievous.

"So what are those two digits?" he asked me.

"Six and a four?"

I was right. The computer was, too. But I was getting tired. My neural nets were misfiring. Another set of numbers flashed.

"How about those?" Hinton said.

"That's tough. Zero and five?" I said.

"Zero and nine. It got zero and nine. It's better than you."

I was wrong. The machine was not.

Paul Voosen is a senior reporter for The Chronicle.