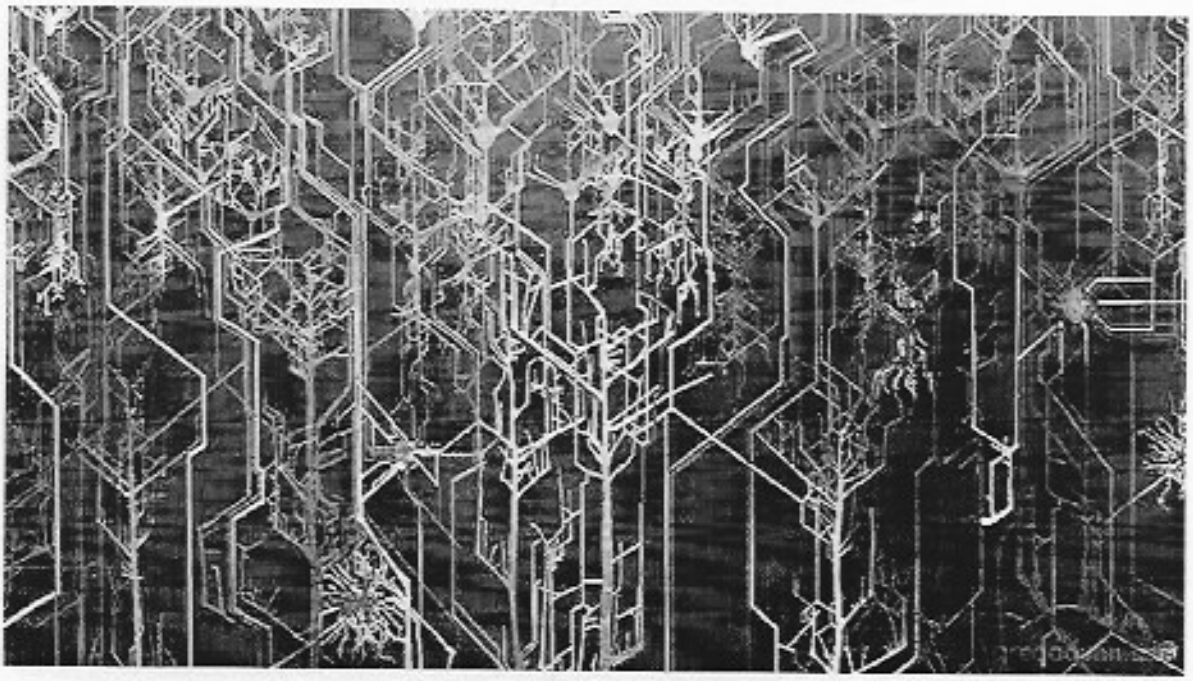
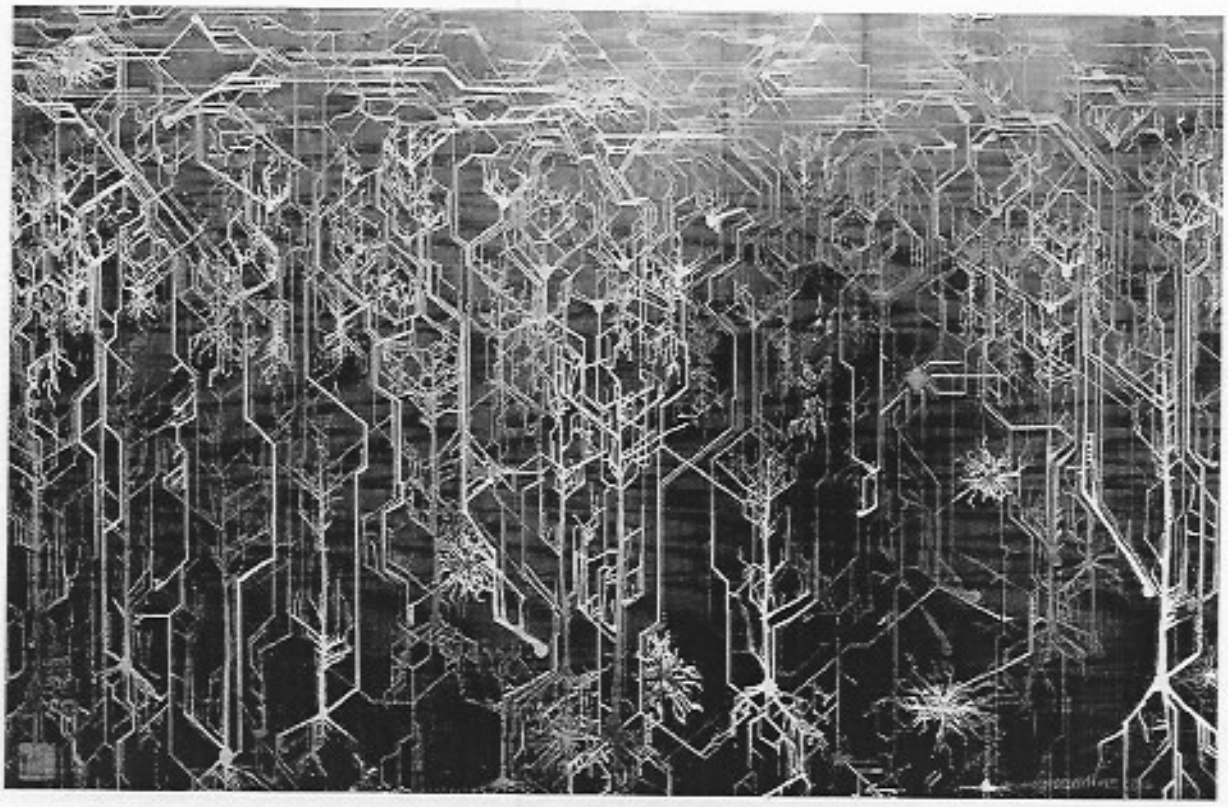
Series: A New Era of Brain-Inspired AI



Mapping the Brain to Build Better Machines

A race to decipher the brain's algorithms could revolutionize machine learning.



Greg Dunn

By Emily Singer

Take a three year-old to the zoo, and she intuitively knows that the long-necked creature nibbling leaves is the same thing as the giraffe in her picture book. That superficially easy feat is in reality quite sophisticated. The cartoon drawing is a frozen silhouette of simple lines, while the living animal is awash in color, texture, movement and light. It can contort into different shapes and looks different from every angle.

Humans excel at this kind of task. We can effortlessly grasp the most important features of an object from just a few examples and apply those features to the unfamiliar. Computers, on the other hand, typically need to sort through a whole database of giraffes, shown in many settings and from different perspectives, to learn to accurately recognize the animal.

Visual identification is one of many arenas where humans beat computers. We're also better at finding relevant information in a flood of data; at solving unstructured problems; and at learning without supervision, as a baby learns about gravity when she plays with blocks. "Humans are much, much better generalists," said Tai Sing Lee, a computer scientist and neuroscientist at Carnegie Mellon University in Pittsburgh. "We are still more flexible in thinking and can anticipate, imagine and create future events."

An ambitious new program, funded by the federal government's intelligence arm, aims to bring artificial intelligence more in line with our own mental powers. Three teams composed of neuroscientists and computer scientists will attempt to figure out how the brain performs these feats of visual identification, then make machines that do the same. "Today's machine learning fails where humans excel," said Jacob Vogelstein, who heads the program at the Intelligence Advanced Research Projects Activity (IARPA). "We want to revolutionize machine learning by reverse engineering the algorithms and computations of the brain."

Time is short. Each team is now modeling a chunk of cortex in unprecedented detail. In conjunction, the teams are developing algorithms based in part on what they learn. By next summer, each of those algorithms will be given an example of a foreign item and then required to pick out instances of it from among thousands of images in an unlabeled database. "It is a very aggressive time-frame," said Christof Koch, president and chief scientific officer of the Allen Institute for Brain Science in Seattle, which is working with one of the teams.

Koch and his colleagues are now creating a complete wiring diagram of a small cube of brain — a million cubic microns, totaling one five-hundredth the volume of a poppy seed. That's orders of magnitude larger than the most-extensive complete wiring map to date, which was published last June and took roughly six years to complete.

By the end of the five-year IARPA project, dubbed Machine Intelligence from Cortical Networks (Microns), researchers aim to map a cubic millimeter of cortex. That tiny portion houses about 100,000 neurons, 3 to 15 million neuronal connections, or synapses, and enough neural wiring to span the width of Manhattan, were it all untangled and laid end-to-end.

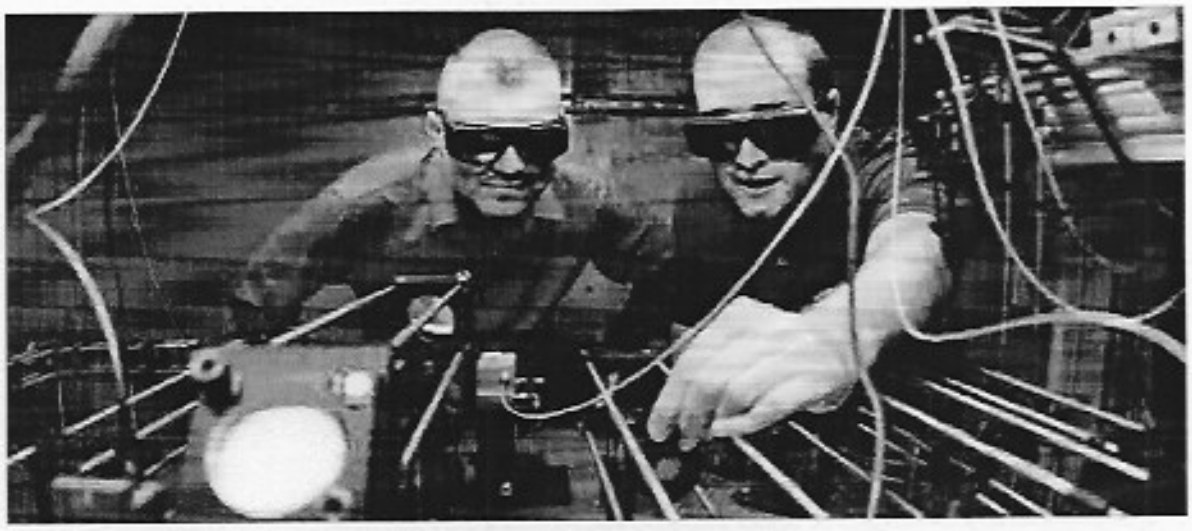
No one has yet attempted to reconstruct a piece of brain at this scale. But smaller-scale efforts have shown that these maps can provide insight into the inner workings of the cortex. In a paper published in the journal *Nature* in March, Wei-Chung Allen Lee — a neuroscientist at Harvard University who is working with Koch's team — and his collaborators mapped out a wiring diagram of 50 neurons and more than 1,000 of their partners. By pairing this map with information about each neuron's job in the brain — some respond to a visual input of vertical bars, for example — they derived a simple rule for how neurons in this part of the cortex are anatomically connected. They

found that neurons with similar functions are more likely to both connect to and make larger connections with each other than they are with other neuron types.

While the implicit goal of the Microns project is technological — IARPA funds research that could eventually lead to data-analysis tools for the intelligence community, among other things — new and profound insights into the brain will have to come first. Andreas Tolias, a neuroscientist at Baylor College of Medicine who is co-leading Koch's team, likens our current knowledge of the cortex to a blurry photograph. He hopes that the unprecedented scale of the Microns project will help sharpen that view, exposing more sophisticated rules that govern our neural circuits. Without knowing all the component parts, he said, "maybe we're missing the beauty of the structure."

The Brain's Processing Units

The convoluted folds covering the brain's surface form the cerebral cortex, a pizza-sized sheet of tissue that's scrunched to fit into our skulls. It is in many ways the brain's microprocessor. The sheet, roughly three millimeters thick, is made up of a series of repeating modules, or microcircuits, similar to the array of logic gates in a computer chip. Each module consists of approximately 100,000 neurons arranged in a complex network of interconnected cells. Evidence suggests that the basic structure of these modules is roughly the same throughout the cortex. However, modules in different brain regions are specialized for specific purposes such as vision, movement or hearing.



Baylor College of Medicine

Andreas Tolias (left), shown here with his student R.J. Cotton, is co-leading one of the Micron teams.

Scientists have only a rough sense of what these modules look like and how they act. They've largely been limited to studying the brain at smaller scales: tens or hundreds of neurons. New technologies designed to trace the shape, activity and connectivity of thousands of neurons are finally allowing researchers to analyze how cells within a module interact with each other; how activity in one part of the system might spark or dampen activity in another part. "For the first time in history, we have the ability to interrogate the modules instead of just guessing at the contents," Vogelstein said. "Different teams have different guesses for what's inside."

The researchers will focus on a part of the cortex that processes vision, a sensory system that neuroscientists have explored intensively and that computer scientists have long striven to emulate.

"Vision seems easy — just open your eyes — but it's hard to teach computers to do the same thing," said David Cox, a neuroscientist at Harvard who leads one of the IARPA teams.

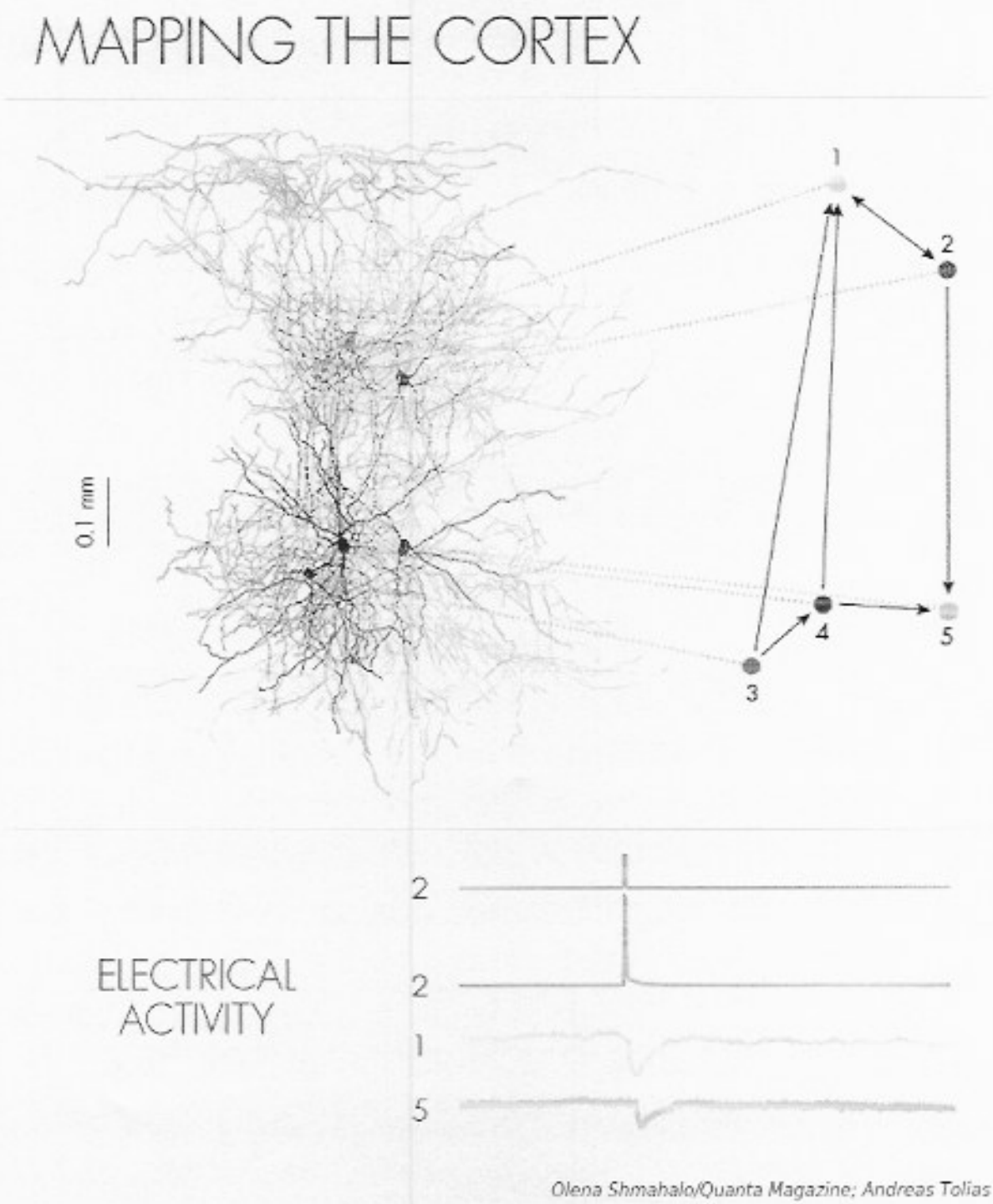
Each team is starting with the same basic idea for how vision works, a decades-old theory known as analysis-by-synthesis. According to this idea, the brain makes predictions about what will happen in the immediate future and then reconciles those predictions with what it sees. The power of this approach lies in its efficiency — it requires less computation than continuously recreating every moment in time.

The brain might execute analysis-by-synthesis any number of different ways, so each team is exploring a different possibility. Cox's team views the brain as a sort of physics engine, with existing physics models that it uses to simulate what the world should look like. Tai Sing Lee's team, co-led by George Church, theorizes that the brain has built a library of parts — bits and pieces of objects and people — and learns rules for how to put those parts together. Leaves, for example, tend to appear on branches. Tolias's group is working on a more data-driven approach, where the brain creates statistical expectations of the world in which it lives. His team will test various hypotheses for how different parts of the circuit learn to communicate.

All three teams will monitor neuronal activity from tens of thousands of neurons in a target cube of brain. Then they'll use different methods to create a wiring diagram of those cells. Cox's team, for example, will slice brain tissue into layers thinner than a human hair and analyze each slice with electron microscopy. The team will then computationally stitch together each cross section to create a densely packed three-dimensional map that charts millions of neural wires on their intricate path through the cortex.

With a map and activity pattern in hand, each team will attempt to tease out some basic rules governing the circuit. They'll then program those rules into a simulation and measure how well the simulation matches a real brain.

Tolias and collaborators already have a taste of what this type of approach can accomplish. In a paper published in *Science* in November, they mapped the connections between 11,000 neuronal pairs, uncovering five new types of neurons in the process. "We still don't have a complete listing of the parts that make up cortex, what the individuals cells look like, how they are connected," said Koch. "That's what [Tolias] has started to do."



Olena Shmahalo/Quanta Magazine; Andreas Tolias

Andreas Tolias and collaborators mapped out the connections among pairs of neurons and recorded their electrical activity. The complex anatomy of five neurons (top-left) can be boiled down to a simple circuit diagram (top-right). Injecting electrical current into neuron 2 makes the neuron fire, triggering electrical changes in the two cells downstream, neurons 1 and 5 (bottom).

Among these thousands of neuronal connections, Tolias's team uncovered three general rules that govern how the cells are connected: Some talk mainly to neurons of their own kind; others avoid their own kind, communicating mostly with other varieties; and a third group talks only to a few other neurons. (Tolias's team defined their cells based on neural anatomy rather than function,

which Wei Lee's team did in their study.) Using just these three wiring rules, the researchers could simulate the circuit fairly accurately. "Now the challenge is to figure out what those wiring rules mean algorithmically," Tolias said. "What kinds of calculations do they do?"

Neural Nets Based on Real Neurons

Brain-like artificial intelligence isn't a new idea. So-called neural networks, which mimic the basic structure of the brain, were extremely popular in the 1980s. But at the time, the field lacked the computing power and training data that the algorithms needed to become really effective. All of the Internet's millions of labeled cat pictures weren't yet available, after all. And although neural networks have enjoyed a major renaissance — the voice- and face-recognition programs that have rapidly become part of our daily lives are based on neural network algorithms, as is AlphaGo, the computer that recently defeated the world's top Go player — the rules that artificial neural networks use to alter their connections are almost certainly different than the ones employed by the brain.

Contemporary neural networks "are based on what we knew about the brain in the 1960s," said Terry Sejnowski, a computational neuroscientist at the Salk Institute in San Diego who developed early neural network algorithms with Geoffrey Hinton, a computer scientist at the University of Toronto. "Our knowledge of how the brain is organized is exploding."

For example, today's neural networks are comprised of a feed-forward architecture, where information flows from input to output through a series of layers. Each layer is trained to recognize certain features, such as an eye or a whisker. That analysis is then fed forward, with each successive layer performing increasingly complex computations on the data. In this way, the program eventually recognizes a series of colored pixels as a cat.

But this feed-forward structure leaves out a vital component of the biological system: feedback, both within individual layers and from higher-order layers to lower-order ones. In the real brain, neurons in one layer of the cortex are connected to their neighbors, as well as to neurons in the layers above and below them, creating an intricate network of loops. "Feedback connections are an incredibly important part of cortical networks," Sejnowski said. "There are as many feedback as feed-forward connections."

Neuroscientists don't yet precisely understand what these feedback loops are doing, though they know they are important for our ability to direct our attention. They help us listen to a voice on the phone while tuning out distracting city sounds, for example. Part of the appeal of the analysis-by-synthesis theory is that it provides a reason for all those recurrent connections. They help the brain compare its predictions with reality.

Microns researchers aim to decipher the rules governing feedback loops — such as which cells these loops connect, what triggers their activity, and how that activity effects the circuit's output — then translate those rules into an algorithm. "What is lacking in a machine right now is imagination and introspection. I believe the feedback circuitry allows us to imagine and introspect at many different levels," Tai Sing Lee said.

Perhaps feedback circuitry will one day endow machines with traits we think of as uniquely human. "If you could implement [feedback circuitry] in a deep network, you could go from a network that has kind of a knee-jerk reaction — give input and get output — to one that's more reflective, that can start thinking about inputs and testing hypotheses," said Sejnowski, who serves as an advisor to President Obama's $100 million BRAIN Initiative, of which the Microns project is a part.

Clues to Consciousness

Like all IARPA programs, the Microns project is high risk. The technologies that researchers need for large-scale mapping of neuronal activity and wiring exist, but no one has applied them at this scale before. One challenge will be dealing with the enormous amounts of data the research produces — 1 to 2 petabytes of data per millimeter cube of brain. The teams will likely need to develop new machine-learning tools to analyze all that data, a rather ironic feedback loop of its own.

It's also unclear whether the lessons learned from a small chunk of brain will prove illustrative of the brain's larger talents. "The brain isn't just a piece of cortex," Sejnowski said. "The brain is hundreds of systems specialized for different functions."

The cortex itself is made up of repeating units that look roughly the same. But other parts of the brain might act quite differently. The reinforcement learning employed in the AlphaGo algorithm, for example, is related to processes that take place in the basal ganglia, part of the brain involved in addiction. "If you want AI that goes beyond simple pattern recognition, you're going to need a lot of different parts," Sejnowksi said.

Should the project succeed, however, it will do more than analyze intelligence data. A successful algorithm will reveal important truths about how the brain makes sense of the world. In particular, it will help confirm that the brain does indeed operate via analysis-by-synthesis — that it compares its own predictions about the world with the incoming data washing through our senses. It will reveal that a key ingredient in the recipe for consciousness is an ever-shifting mixture of imagination plus perception. "It is imagination that allows us to predict future events and use that to guide our actions," Tai Sing Lee said. By building machines that think, these researchers hope to reveal the secrets of thought itself.