COMPUTATIONAL NEUROSCIENCE

A Romance Blossoms Between Gray Matter and Silicon

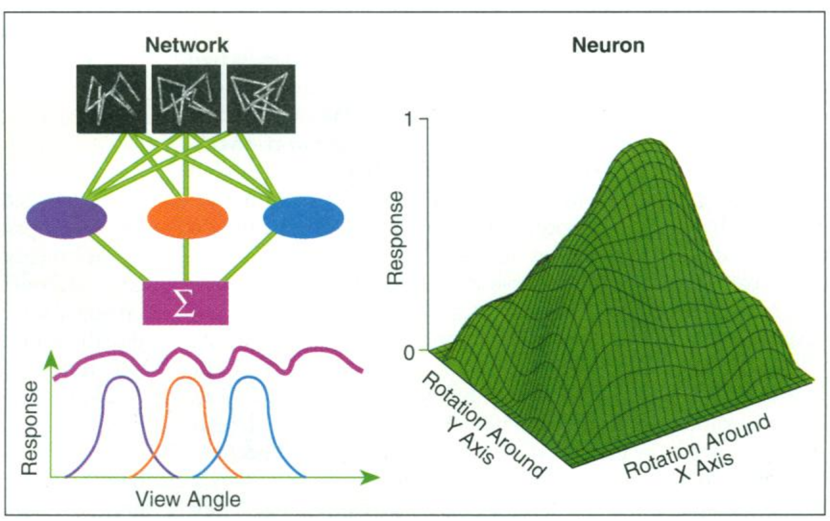
To a computer, few things are more difficult than recognizing a three-dimensional object. Just change the viewing angle, and a well-known object can look completely unfamiliar: Think of a fork seen from the end. Finding a way for a machine to recognize an object without giving it a bulging file of images showing the object from every possible angle is one of the toughest problems in machine vision, but one team, working with a simple mathematical model of vision, may have taken a step toward solving it. Neuroscientists face a related challenge—explaining how the brain effortlessly manages the trick. There too, a research group recently took a step forward, by monitoring patterns of brain activity in monkeys. By a remarkable coincidence, both the machine-vision researchers and the brain specialists have ended up focusing on exactly the same theory.

Well, maybe it's not such a coincidence, as the two teams are actually one and the same. A collaboration led by Thomas Poggio at the Massachusetts Institute of Technology and Nikos Logothetis at the Baylor College of Medicine, the team is just one of many taking part in a budding scientific romance. After three decades of mutual distrust, neuroscientists and artificial intelligence (AI) researchers are starting to find that there's wide, profitable overlap between the two fields. "There's a real convergence now between neuroscience and what people are actually building," says Paul Werbos, program director for neuroengineering at the National Science Foundation. "These tools have brought us to the verge of Newtonian revolution in brain science."

The value of such collaboration now seems obvious. But for a long time the tacks taken by researchers in the two fields could hardly have been more different. Artificial intelligence researchers traditionally ignored how the brain actually works, focusing instead on abstract principles behind the highest levels of thought, such as mathematical reasoning and language ability, so that those principles could be embodied in computers. Neuroscientists, on the other hand, often shied away from thinking about the brain as an information-processing system and concentrated on anatomy and biochemistry that could only vaguely be linked to the complex behavior of people and animals.

By now, however, neuroscientists have come so far in picking apart some of the brain's key physical structures that many feel it is time to take the next step: determining how these structures produce thought. "Neuroscientists have traditionally been engaged in decomposing the brain into its component parts," says Terrence Sejnowski, a Howard Hughes Medical Institute researcher who is also on the faculty at the University of California, San Diego, and the Salk Institute. "Modeling on a computer is a way of putting the pieces back together."

Meanwhile, AI researchers relying on the traditional effort to identify higher principles in thought haven't made much progress. The brain-inspired programs known as "artificial neural networks," or ANNs, though, are another story. ANNs simulate, usually in software, what the brain provides in hardware: a vast array of interconnected switches—neurons—that manipulates information in the form of a complex, ever-changing pattern of signals. ANNs have made great strides in their ability to recognize rudimentary images and other types of patterns, a task at which the brain does superbly and conventional AI programs fail miserably. Now, many ANN researchers have concluded that vaguely modeling these programs on the brain's extreme connectivity isn't enough; if ANNs are to perform anywhere near the brain's level, they must be made to resemble more closely the brain's logical and even physical structure.



Photographic memory. Individual "neurons" in a neural network respond to specific snapshots of a 3D object, but the network as a whole can recognize it in any orientation. A monkey neuron shows a similar preference for a specific orientation, suggesting that the brain may use the same scheme as it learns to recognize shapes.

SOURCE: THOMAS POGGIO AND NIKOS LOGOTHETIS

Lessons from the brain. To that end, many researchers interested in building better functioning ANNs have begun working in close collaboration with neuroscientists. Take Carnegie-Mellon University computer scientist James McClelland, who is intrigued by the widely held theory that the brain possesses two different learning systems. One, which appears to be centered in a structure called the hippocampus, seems to specialize in quickly learning and retaining information such as the names and records of players on a baseball team—without doing anything much with those data. A second system, probably in the neocortex, is better suited to taking specific data and gradually extracting generalized information from them, such as the insight that poor relief pitching tends to catch up with a team sooner or later.

Both types of information processing are important: Without gradual generalization, we wouldn't be able to learn to drive a car; without quick learning, we wouldn't be able to remember where we parked it. McClelland believes ANNs require a similar setup—and for much the same reason. A handwriting-recognizing ANN might, for example, need a hippocampuslike memory to learn specific individuals' handwriting, and a neocortexlike module to generalize the information and learn how to recognize anybody's handwriting. "Most neural-network models tend to do one or the other well, but not both," he says.

With University of Arizona neuroscientist Bruce McNaughton, McClelland came up with a computer model that imitates the two learning systems they think the brain employs. In this model, specific patterns are quickly learned and stored, hippocampus-style. Then these patterns are replayed over and over again into a separate ANN that, like the neocortex, allows new data to subtly and gradually modify old data. In effect, it constructs a generalization from many specific data.

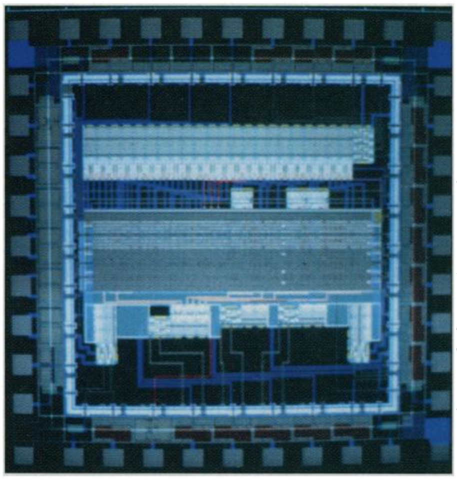
Werbos, too, is keeping his eye on how the brain learns as he builds ANNs. He's trying to mimic two of the brain's most powerful learning aids. One is emotion, which leads us to identify and carry out actions associated with good feelings while avoiding those that make us unhappy. The other is the ability to form expectations—to recognize patterns that foreshadow what is going to happen next. Because they accelerate learning, Werbos contends, both features are vital for ANNs that would perform complex functions such as controlling dangerous machinery. An "emotional" ANN, he explains, could be programmed to regard some real-world parameter, such as keeping a nuclear reactor core at a safe temperature, as a form of "happiness," in the sense that all its actions will be aimed at maximizing this parameter. And if the ANN could also predict a rise in core temperature based on coolant temperature and flow rates, then take early action, it would be that much better equipped to achieve emotional satisfaction.

Commercial ANN chips based on the kinds of designs Werbos and colleagues have pioneered are already being tested by Ford, Mercedes-Benz, and other carmakers. The chips would monitor a car's engine to save fuel or detect potential crashes and engage alarms or air bags. And a new version of the F-15 fighter jet goes into testing in September with chips that employ a version of emotion-motivated learning to back up the pilot if the plane is hit. "A neural network might learn faster than a human pilot can to control a plane that suddenly behaves differently than the way it was designed," explains Werbos.

Working just at the other side of the ANN-neuroscience boundary are researchers interested in what ANNs can tell them about the brain—even going so far as to build ANNs, then partially disconnect them to simulate brain damage. Brain-damaged patients often lose what appear to be very specific capabilities, such as the ability to use certain parts of speech, and researchers have generally assumed that the lesion indicates a brain region dedicated to the lost capability. But by building an ANN and disconnecting a region of it to simulate the lesion suffered by a patient, University of Pennsylvania cognitive neuroscientist Martha Farah has shown that the assumption may not be valid.

Farah's human patient primarily showed an inability to recall most knowledge about living things, such as the fact that cows have udders. But because the patient retained a small amount of knowledge about living things and lost a small amount of other types of knowledge, Farah came to suspect that the damaged area actually contributed to a more basic function: visual memory processing. Knowledge about living things tends to be visual, she reasoned; you picture an udder if asked if cows have them, but you probably don't pull up a strong image if someone asks you the capital of New Jersey. Sure enough, the ability of Farah's ANN to assemble most, but not all, memories of living things deteriorated when she disconnected a section of its visual memory area, and other types of knowledge were only slightly affected.

The result suggests that a patient's observable impairment doesn't always map in an obvious way to the function of the damaged part of the brain, says Farah. "Neuropsychologists have avoided even thinking about this possibility, because how could you trace function in the huge interactivity of the brain?" she adds. But having raised that possibility, ANNs also give neuroscientists a tool for understanding more complex relations between anatomy and function. As Farah puts it, "The computer gives you a great way of modeling what happens when you take a piece of that interactivity away."



RODNEY DOUGLAS AND MISHA MAHOWALD/ MRC ANU/OXFORD

Silicon neurons. A prototype chip, 1.2 millimeters across, includes five "neurons" that simulate ion currents and synaptic responses.

Walking the line. Most of the researchers involved in this ongoing merger work on one side of the fence or the other, but a small but growing number manages to make important contributions to both AI and neuroscience at the same time. Until recently, such attempts earned a cold shoulder from both fields, notes Boston University's Stephen Grossberg, one of the pioneers of the dual-track approach. "Now, thank God, the prejudices are disappearing," says Grossberg, whose brain models and ANNs, once largely ignored, have come to be widely respected.

Poggio and Logothetis are beneficiaries of this more hospitable environment. Their laboratories studied the three-dimensional object-recognition problem on an ANN and in monkeys. One way a brain or computer might recognize an object from any angle is to construct a complete three-dimensional image, then rotate the image until it matches the view to be recognized. But Poggio felt that, for both brains and computers, such an approach would take far too much computing power and memory.

His hypothesis: The brain remembers a small number of two-dimensional "snapshots" of the object, each from a different angle. When shown a view of the object, the neurons dedicated to each snapshot fire at a rate proportional to how closely the view matches that snapshot. Even if the view presented doesn't exactly match any of them, it should at least get a few neurons to fire weakly, and that should be enough to trigger a match.

Working with ANNs, Poggio and his colleagues were able to show that the scheme can work. They built a computer system that maintains as few as four snapshots but can reliably recognize certain three-dimensional objects from any angle. In the last few months Logothetis has gathered evidence that monkeys, at least, actually use this scheme. He and his colleagues recorded the activity of individual neurons while the animals viewed familiar and unfamiliar objects at various orientations; they found that some neurons in the visual cortex do seem to be dedicated to certain snapshots of an image. The suspect neurons fired strongly for a familiar object at a particular angle—presumably one that matched the snapshot—weakly for the same object at other angles, and not at all for the wrong object. "The monkey data correspond to the model so closely it's almost too good to be true," says Poggio.

These results suggest some of the satisfactions that researchers are finding at the interface of brain and machine. Yet those same researchers also face one persistent frustration: When it comes to neural-network computing, digital hardware is a poor substitute for the brain's "wetware." A digital circuit can express information only in discrete chunks, whereas neurons continuously vary their firing rate to express a much richer range of information at higher speed. To narrow the gap, Oxford University neuroscientist Rodney Douglas and his colleague Misha Mahowald have been working on "silicon neuron" chips—chips based on the same microelectronic technology as ordinary computer chips but relying on analog rather than digital signals. Douglas's chips can already imitate the widely varying electrical signals with which neurons respond to different inputs.

Now he and Mahowald are trying to combine the silicon neurons into neural networks, so that they and other neuroscientists will have an even more faithful replica of the brain on which to conduct experiments that shed light on the brain's functioning. Douglas even reluctantly admits to hoping that a "brain" made of silicon neurons might someday rival the real thing. "We're always cautious about thinking along those lines," he says, "but of course that's what we're trying to do."

Perhaps it won't be all that long before scientists begin to wonder why anyone ever bothered to distinguish artificial intelligence from the "real" thing.

–David H. Freedman

David H. Freedman's book about artificial intelligence, Brainmakers, *was recently published by Simon and Schuster.*