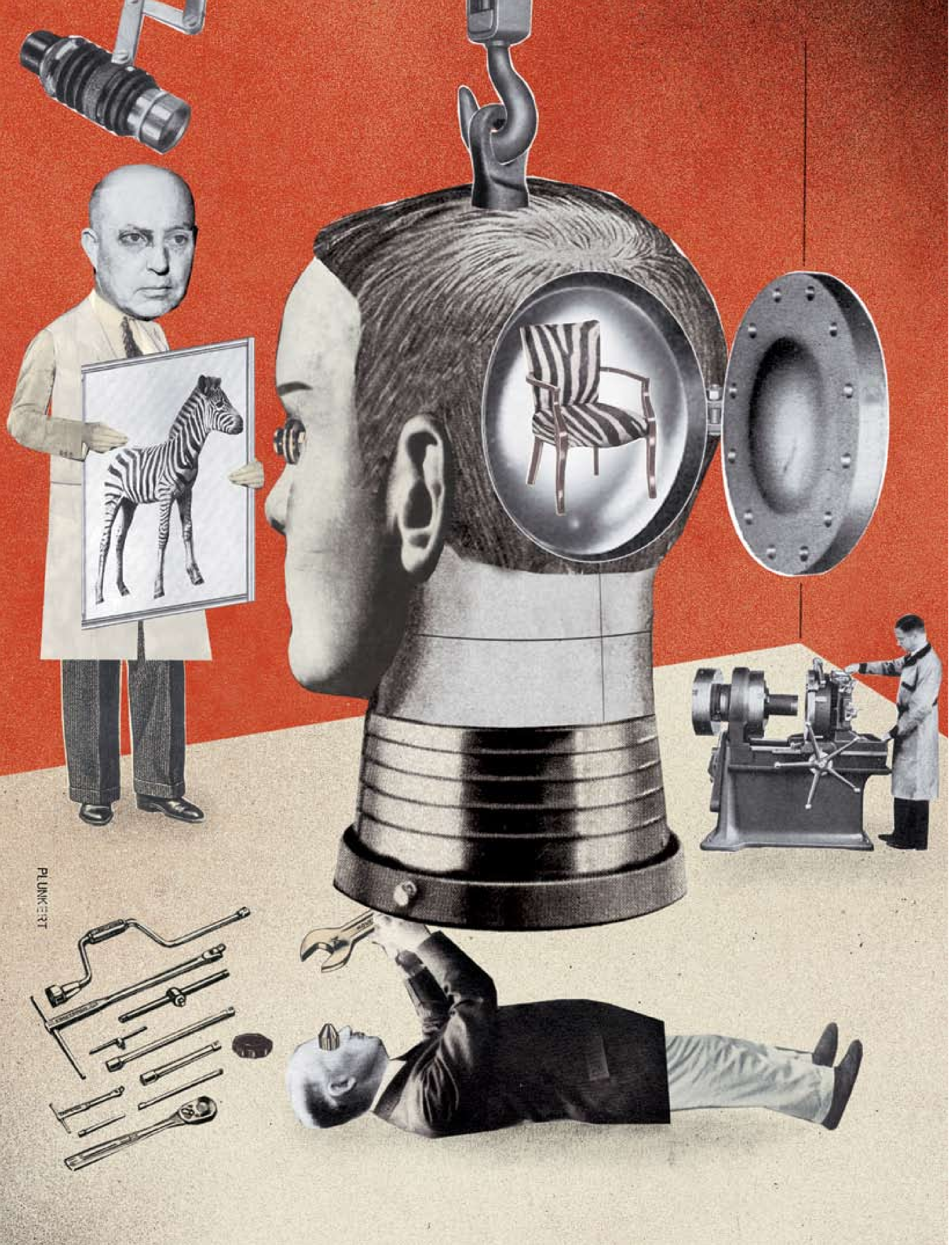


Reverse-Engineering the Brain

At MIT, neuroscience and artificial intelligence are beginning to intersect.

By Fred Hapgood
Illustration by David Plunkert

"Maggie is a *very* smart monkey," says Tim Buschman, a graduate student in Professor Earl Miller's neuroscience lab. Maggie isn't visible—she's in a biosafety enclosure meant to protect her from human germs—but the signs of her intelligence flow over two monitors in front of Buschman. For the last seven years, Maggie has "worked" for MIT's Department of Brain and Cognitive Sciences (BCS). Three hours a day, the macaque plays computer games that (usually) are designed to require her to generate abstract representations and then use those abstractions as tools. "Even I have trouble with this one," Buschman says, nodding at a game that involves classifying logical operations. But Maggie is on a roll, slamming through problems, taking about half a second for each and getting about four out of five right.

Maggie's gaming lies at the intersection of artificial intelligence (AI) and neuroscience. Under the tutelage of Buschman and Michelle Machon, another graduate student, she is contributing to research on how the brain learns and constructs logical rules, and how its performance of those tasks compares with that of the artificial neural networks used in AI.

Forty years ago, the idea that neuroscience and AI might converge in labs like Miller's would have been all but unthinkable. Back then, the two disciplines operated at arm's length. While neuroscience focused on uncovering and describing the details of neuroanatomy and neural activity, AI was trying to develop an independent, nonbiological path to intelligence. (Historically, technology hasn't really needed to copy nature that slavishly; airplanes don't fly like birds and cars don't run like horses.) And it was AI that seemed to be advancing much more rapidly. Neuroscience knew hardly anything about what the brain was, let alone how it worked, whereas everyone with an ounce of sense believed that the day when computers would be able to do everything humans did (and do it better) was well within sight. In 1962, President Kennedy himself was persuaded of the point, pronouncing automation (or as it was often called then, "cybernation") the core domestic challenge of the 1960s, because of the threat that it would put humans out of work.

But something derailed the AI express. Although computers could be made to handle simple objects in a controlled setting, they failed miserably at recognizing complex objects in the natural world. A microphone could distinguish sound levels but not summarize what had been said; a manipulator could pick up a clean new object lying in an ordered array but not a dirty old one lying in a jumbled heap. (Nor could it, in Marvin Minsky's inspired example, put a pillow in a pillowcase.) Today we worry far more about competition from humans overseas than about competition from machines.

While AI's progress has been slower than expected, neuroscience has gotten much more sophisticated in its understanding of how the brain works. Nowhere is this more obvious than in the 37 labs of MIT's BCS Complex. Groups here are charting the neural pathways of most of the higher cognitive functions (and their disorders), including learning, memory, the organization of complex sequential behaviors, the formation and storage of habits, mental imagery, number management and control, goal definition and planning, the processing of concepts and beliefs, and the ability to understand what others are thinking. The potential impact of this research could be enormous. Discovering how the brain works—*exactly* how it works, the way we know how a motor works—would rewrite almost every text in the library. Just for starters, it would revolutionize criminal justice, education, marketing, parenting, and the treatment of mental dysfunctions of every kind. (Earl Miller is hoping the research done in his lab will aid in the development of therapies for learning disorders.)

Such progress is one reason the once bright line between neuroscience and AI is beginning to blur at MIT—and not just in Miller's lab. Vision research under way at the Institute also illustrates how the two disciplines are beginning to collaborate. "The fields grew up separately," says James DiCarlo, assistant professor of neuroscience, "but they're not going to be separate much longer." These days, AI researchers follow the advance of neuroscience with great interest, and the idea of reverse-engineering the brain is no longer as implausible as it once seemed.

Understanding Object Recognition

Much of the work in DiCarlo's lab concerns object recognition, which is what allows us to identify an object (such as a cow) in many different presentations (cows far away, cows viewed from above, cows at dawn, a cow in a truck) without mistaking it for similar objects (like, say, a horse). DiCarlo and graduate student David Cox published research last August in *Nature Neuroscience* that focused on one of the basic questions about object recognition: how much of our success in recognizing objects depends on hard-wired, innate circuitry, and how much on learned skills?

DiCarlo and Cox conducted each of their experiments on a dozen people, one person at a time. Subjects sat in front of equipment that could both display images of objects and track the direction of the subjects' gaze. The objects were computer generated and looked vaguely like anthropomorphized animals, but they were designed to be unfamiliar to the subjects. An object would appear in one of three positions on a screen, and the subject would naturally shift his or her gaze toward it. For certain objects, however, the

James DiCarlo's lab has shown that visual experience affects the ability to recognize objects.

KENT DALTON

researchers would substitute new objects while the subjects were moving their eyes. For example, let's say an object that looked kind of squat, with perky ears, was introduced at the right of the screen while the subject was focusing on the center. As the subject's gaze shifted toward squat and perky, the researchers would replace the object with one that looked slightly thinner, with droopier ears. Since humans are effectively blind during gaze shifts, the subjects did not notice the swap. But their brains did.

After an hour or two of exposure to different objects, some of which were consistently swapped out when they appeared in particular positions, subjects were presented with pairs of the objects in different positions on the screen and asked to compare them. One might expect that the subjects would distinguish the objects without much difficulty. And so they did, except when the objects had been swapped—and were now reappearing in the same positions where the swaps occurred. Subjects tended to confuse those objects: that is, they were more likely to judge that squat and perky at one position and thin and droopy at another were one and the same object. DiCarlo thinks such errors show that the brain's mechanisms for recognizing the same object in different places depend on normal visual experience across space and time. "The finding suggests that even fundamental properties of object recognition may be developed through visual experience with our world," he says. DiCarlo and his team are conducting similar experiments in animals to examine the patterns of neuronal activity that underlie object recognition. (A good example of this research was published in the November 4, 2005, issue of *Science* magazine. DiCarlo and three collaborators recorded and analyzed the activity of hundreds of neurons in macaque brains. They were able to show that highly reliable information about object identity and category was contained in even handfuls of neurons.)

Discovering how the brain works—*exactly* how it works, the way we know how a motor works—would rewrite almost every text in the library.

Object recognition has been one of the major targets, and major disappointments, of traditional AI. While machine vision is a real industry, its successes have been in narrowly defined applications under highly controlled conditions, such as decoding license plates, identifying fingerprints, recognizing printed characters, and inspecting products (for instance, identifying burnt potato chips so they can be blown out of an assembly line). Each machine vision system "sees" only a specific kind of object; for example, the machine that reads license plates would not be able to identify fingerprints, and vice versa. Although today's technology might be good enough to give us machines that recognize any one thing, most jobs in most industries—assembly, maintenance, health care, transportation, security—require more versatility than that. Workers need to be able to recognize a hammer and a screwdriver and a wrench, despite differences in lighting, the objects' orientation, and the surrounding clutter. The failure to build machines that can do this is especially frustrating given that birds like crows, and small mammals like rats, routinely exhibit a level of skill in general recognition that is way beyond current technology. There is something about not being able to make machines as smart as we are that is consoling to our vanity; but not being able to make one as smart as a pigeon is just embarrassing.

So for years AI researchers have been working on the problem of associating visual patterns with meanings or identities. This is one of the areas where AI and neuroscience have been edging toward each other: neuroscience has been working on the brain's role in object recognition, AI on the general logic of what any system would have to do to solve the same problem. After decades they are almost within talking distance. DiCarlo wonders if it might be time to christen a new discipline that draws from both fields, like "biologically inspired machine vision."

No university is approaching this intersection faster than MIT, where the collaboration of engineering and science is an institutional mission. And that, says DiCarlo, is one reason he came to MIT: he expects the revolution to happen here.

Modeling Immediate Recognition

A striking illustration of DiCarlo's point can be found in the labs of Tomaso Poggio. The codirector of MIT's Center for Biological and Computational Learning, Poggio has been working on vision for four decades, first at the Max Planck Institute in Tübingen, Germany, then at MIT's AI lab (which became the Computer Science and Artificial Intelligence Lab), and now in the Department of Brain and Cognitive Sciences. (Poggio collaborated with DiCarlo in the macaque experiments described in *Science*.) For much of this time, Poggio directed one research group in neuroscience and one in machine vision and saw no reason to bring them together. "We knew so little," he says. "I always thought it was a mistake to expect much from neuroscience." But recent results from a project carried out by postdoc Thomas Serre and Aude Oliva, assistant professor of cognitive neuroscience in BCS, made him a convert.

Poggio's lab is currently focusing on a type of object recognition called immediate recognition. This phenomenon was first described in 1969 in a paper by MIT lecturer

Mary Potter (now a professor of psychology at BCS) and her research assistant, Ellen Levy. Immediate recognition is the fastest known form of recognition. A subject in a classic immediate-recognition experiment is seated before a display and asked to push one of two keys in response to each image in a series, depending on whether it contains an animal or not. To make sure looking at one image doesn't accidentally help subjects learn how to look at others, researchers choose images that are very different: many species, in many different poses and perspectives, set against a wide range of backgrounds. The photos come and go in a few tenths of a second. At the start of a study, a subject might have next to no awareness of even being shown an image, let alone recognizing what is in it. Yet amazingly, people hit the right keys more often than not. They get steadily better—and become conscious of the appearance of the images—with practice. Still, at the outset, something in the brain is able to recognize and categorize objects before the subject is even aware of seeing anything.

"We knew so little. I always thought it was a mistake to expect much from neuroscience." –Tomaso Poggio

Immediate recognition is important to researchers because it is the simplest possible case of general object recognition. It happens too quickly to involve recruiting lots of neurons or processing information intensively or sending and receiving impulses over more than a fraction of a centimeter. Information from eye movements, a key element in other kinds of recognition (as in DiCarlo's work), can play no role. Yet somehow the right keys get pressed (mostly), which means that a limited form of general-purpose object recognition must be possible using a relatively small number of neurons organized in a relatively simple fashion.

Building on work Poggio did with Max Riesenhuber, PhD '00, then a grad student at MIT and now a professor at Georgetown University, Serre, Poggio, and others in Poggio's group developed a theory about the part of the visual cortex chiefly responsible for immediate recognition. Their approach to visual processing was in many respects different from a machine vision engineer's. For instance, most machine vision programs feature one processor executing a series of instructions in consecutive order, an architecture known as "serial processing." The brain, on the other hand, uses "parallel processing," an approach in which a problem is broken up into many pieces, each tackled separately by its own processor, after which the results are combined or integrated to get a single general result—say, the perception of a cow. In theory, engineers could use parallel processing for machine vision programs (and some have tried), but in practice it is seldom obvious how to break down a problem in a way that allows the finished pieces to be seamlessly recombined.

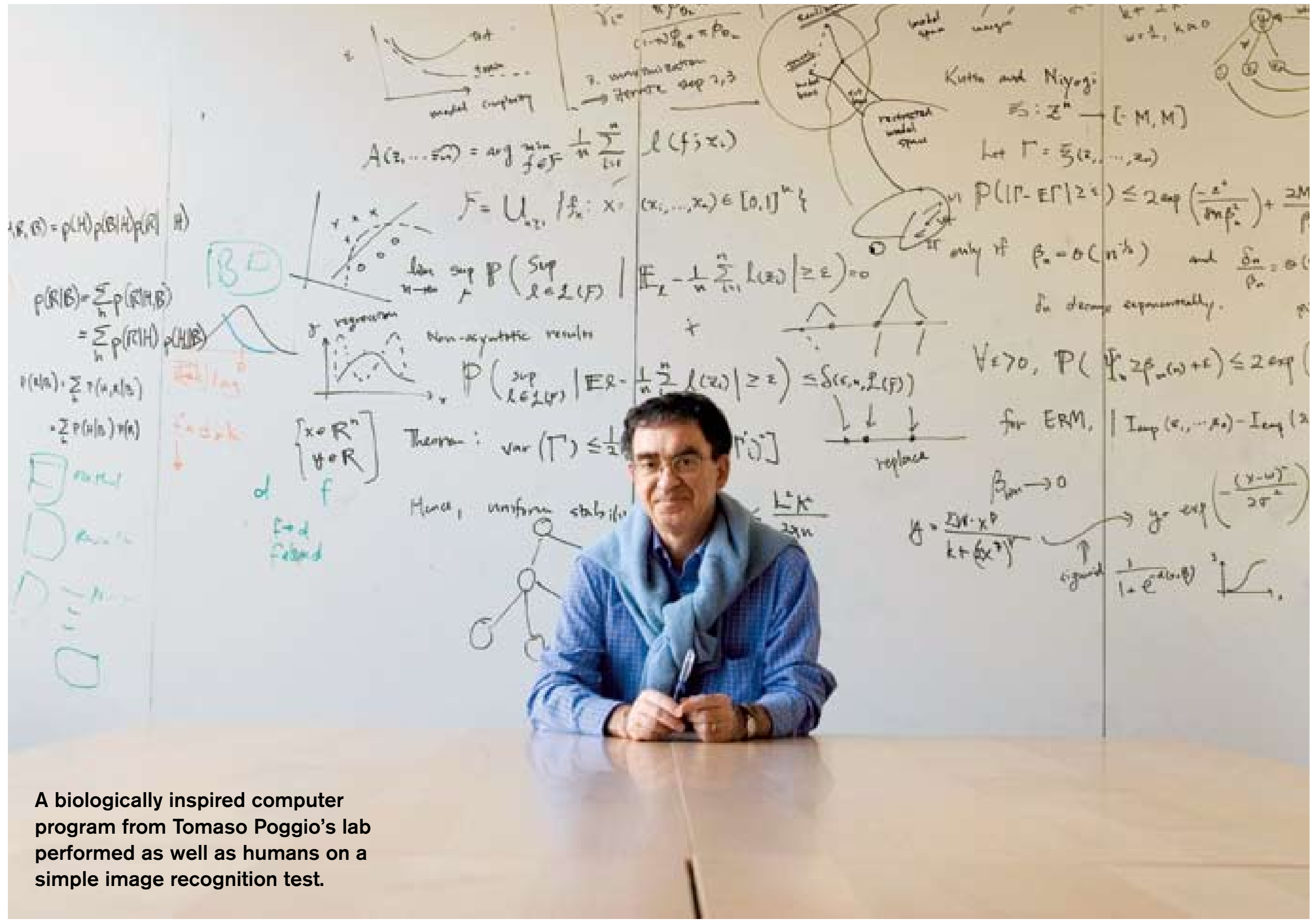
Biological vision solves this problem in several different ways. One, according to Poggio's group, is to organize processing around two simple operations and then alternate these operations in an ordered way through layers of neurons. Layer A might filter the basic inputs from the optic nerve; layer B would integrate the results from many cells in layer A; C would filter the inputs from B; D would integrate the results from C; and so on, perhaps a dozen times. As a signal rises through the layers, the outputs of the parallelized processors gradually combine, identity emerges, and noise falls away.

Serre and Poggio used this layering technique to enable their model to do parallel processing. Another trick they borrowed from biology was to increase the number of connections linking their basic switching units. The switching units in conventional computers have very few connections, usually around three; neurons, the basic switching units of the brain, have thousands or even tens of thousands. Serre and Poggio endowed the logical switches in their model with a biologically plausible degree of connectivity. In cases where the science was not yet known, they made assumptions based on their broader experience with neuroanatomy.

To test their theory, Serre and Poggio developed an immediate-recognition computer program that analyzes digital images. When digital image files are fed into the program, it passes them through multiple alternating layers of filtering and integrating cells, training itself to identify and classify the images. "The key is building complexity slowly," Serre says. "Introducing intelligence too quickly is a big mistake." Early AI efforts may have tried to zero in on identity too quickly, throwing out information that was critical for getting the right answer.

Serre and Poggio's approach was a spectacular success. From a neuroscientific point of view, some of their assumptions turned out to predict real features, such as the presence of cells (call them OR cells) that pick the strongest or most consistent signal out of a group of inputs and copy it to their own output fibers. (Imagine a group of three neurons, A, B, and C, all sending signals to OR neuron X. If those signals were at strength levels 1, 2, and 3 respectively, X would suppress A and B and copy C's signal to its output. If the strengths had been 3, 2, and 1, it would instead have copied A's signal and suppressed those of B and C.)

The results were just as dramatic from an AI point of view. When human subjects and Serre and Poggio's immediate-recognition program took the animal presence/absence test,



A biologically inspired computer program from Tomaso Poggio's lab performed as well as humans on a simple image recognition test.

KENT DALTON

the computer did as well as the humans—and better than the best machine vision programs available. (Indeed, it got the right answer 82 percent of the time, while the humans averaged just 80 percent.) This is almost certainly the first time a *general*-vision program has performed as well as humans.

The promising results have Poggio and Serre thinking beyond immediate recognition. Poggio suspects that the model might apply just as well to auditory perception. Serre advances an even more daring speculation: that general object recognition is the basic building block of cognition. Perhaps that's why we say "I see" when we want to indicate that we understand something.

Although extending their theory in these new directions will take some work, Serre and Poggio's model has already begun to spread through both the AI and neuroscience communities at MIT. Electrical-engineering graduate student Stan Bileschi recently finished a doctorate that applied the model to scene recognition, which is the derivation of higher-order judgments—"it's a farm!"—from the recognition of separate objects—a barn, a cow, a split-rail fence. Bileschi believes that general scene analysis will be critical to many real-world machine vision applications—surveillance, for instance.

Immediate recognition is the foundation of overall visual recognition, says Poggio, but it's not all there is to it. There are many levels of recognition, and immediate recognition is one of the simplest. Depending on the context, an object might be identified as a toy, a doll, a Barbie, a reflection of American culture, a female, a representation of a girl with a weird growth disorder, and so on, down a long list. Similarly, in chess problems, recognizing the right move can take seconds or minutes or hours, depending on the configuration of the pieces. Presumably, as problems get harder, solving them requires recruiting higher levels of brain function—and that takes time.

An immediate-recognition model might solve the vision problems that have impeded the development of useful maintenance and construction robots. Or we might find that to be really useful, such robots need to be able to recognize both anomalies in the landscape and their causes. That type of recognition is clearly of a higher order.

The next step is to build recognition models that recruit more and more resources, and thus require more processing time. "We know how the model could be changed to include time," says Serre. "This might bring us closer to thinking—just maybe." ■