A reprint from

American Scientist

the magazine of Sigma Xi, The Scientific Research Society

What's in Brian's Brain?

Brian Hayes

OF ALL THE COMPUTING devices I encounter from day to day, the most mysterious is the one in my own head. Other machines—from gadgets in my pocket to unseen Internet servers—process information in ways that I think I understand. When it comes to the brain, however, I haven't got a clue.

Filling in the blanks in our knowledge of the brain is the mission of several ongoing and recently announced research programs. The Allen Institute for Brain Science in Seattle has been compiling atlases of nerve tissue. At Harvard University the Connectome Project is tracing the links between individual nerve cells. The similarly named Human Connectome Project, a collaboration of four institutions, looks at the brain's wiring diagram on a larger scale.

"Functional connectomics" would go a step beyond circuit diagrams to active wiretapping—listening in on the signals that pass through all the nerve fibers in a network. A project of this kind called the Brain Activity Map has become the centerpiece of a program announced in April by President Barack Obama, with the promise of $100 million in first-year federal funding.

Meanwhile the European Commission has committed €1 billion ($1.3 billion) over 10 years to the Human Brain Project, which emphasizes computational models of nervous-system activity. It will build on an existing simulation called the Blue Brain Project.

The questions addressed by these endeavors are so intriguing—and our present ignorance of how the brain works is so irksome—that I find myself rooting enthusiastically for the pro-

Despite the progress of neuroscience, I still don't know my own mind

grams' success. Yet some of the questions are also so difficult that it's hard to set aside skepticism.

What's the Big Idea?

Advocates of the new brain science initiatives cite the Human Genome Project as a precedent. But the comparison between neuroscience and molecular biology is more illuminating if we take a broader historical view.

A century ago, biologists knew next to nothing about the molecular basis of genetics and metabolism. Some of what they thought they knew was wrong: As late as the 1940s the stuff of genes was believed to be protein, not DNA. Then, over a span of a dozen years, the central mechanisms of life were suddenly revealed in vivid detail. The double helix and the genetic code provided the key to understanding both inheritance and the control of chemical synthesis within the cell. The essential idea was surprisingly simple: For many purposes one could ignore all the biochemical details and look upon genetic information as an abstract sequence of symbols, a message written in the four-letter alphabet of DNA. Without that level of abstraction, the Human Genome Project would have been unthinkable.

Neuroscience has followed a different trajectory. Early in the 20th century, knowledge of neural anatomy and physiology was already quite advanced. The main elements of the nervous system were recognized as indi-

vidual cells (neurons) with input fibers (dendrites) and output fibers (axons). Stimuli reaching the dendrites cause the cell to "fire," producing an impulse, or "spike," on the axon. The physics of the nerve impulse was also understood: Ions flow across the cell membrane, creating an electrical disturbance that propagates as a wave along the fiber. One cell communicates with the next through a synapse, where axon and dendrite are pressed together. In the 1940s Warren S. McCulloch and Walter H. Pitts showed that small networks of neurons could implement basic logic functions. And then Donald O. Hebb proposed a mechanism of learning and memory in which neurons that frequently fire together develop stronger synaptic ties.

At mid-century, neuroscience seemed poised for a breakthrough. And indeed there were dozens of momentous discoveries—a torrent of new knowledge about the detailed structure and function of nervous tissue. What *hasn't* emerged is a big idea with the explanatory power of the double helix or the genetic code. We still can't read out the information stored or embodied in a brain—the skills an organism has acquired, the facts learned, the experiences remembered—as we can read out information encoded in a strand of DNA. None of the pending brain study projects have promised to supply such a mind-reading capability. But perhaps they will at least offer some hints about how information is represented and stored in the brain.

The Brainiac

As modern neuroscience has grown up alongside molecular biology, it has also been strongly influenced by another blossoming field: computer science. Brains and computers are called on to perform many of the same tasks: doing arithmetic, solving puzzles, planning strategies, analyzing pat-

terns, filing away facts for future reference. At a more theoretical level, biological and electronic machines surely have the same computational power. And yet attempts to explain brains in terms of computers, or vice versa, have not been wildly successful.

The neural circuits sketched by McCulloch and Pitts look much like the diagrams of logic gates (AND, OR, NOT, etc.) that appear in designs of digital computers. The resemblance is misleading. McCulloch and Pitts showed that their networks of neurons can compute the same set of logical propositions as certain abstract automata. However, neurons in the brain are not wired together to form such simple logic circuits.

A typical electronic logic gate computes a function of two or three inputs. For example, a three-input AND gate has an output of *true* if and only if all three inputs are *true*. There are eight combinations of three *true*-or-*false* signals, yielding 256 possible functions of those inputs. A typical neuron has thousands of synaptic inputs. For 1,000 signals there are $2^{1,000}$ combinations—an immense quantity, far exceeding the number of cells in the human brain. The number of functions that a 1,000-input neuron might be computing is larger still: 2 raised to the power $2^{1,000}$. For such a neuron there's no hope of constructing a complete truth table, showing the cell's response to all possible combinations of inputs; the appropriate tool for describing the action of the neuron is not logic but statistics.

The spiky nature of neural signaling is a further complication. Standard electronic logic gates work on persistent signals: Apply a voltage to each of the inputs, wait for the system to settle down, then read the state of the output. Neural signals are brief impulses rather than steady voltages. Thus the output of a neuron depends not just on which input signals are present but also on their precise timing.

Abstract models of the brain, such as artificial neural networks, smooth away many of these complexities. If the function that a neuron computes is just a weighted sum of the inputs, only a tiny fraction of the possible combinations need to be considered. (For a fixed set of input weights, the number of distinguishable combinations falls from $2^{1,000}$ to 1,001.) Likewise the problem of synchronizing spikes is eased by assuming that the neuron merely measures the
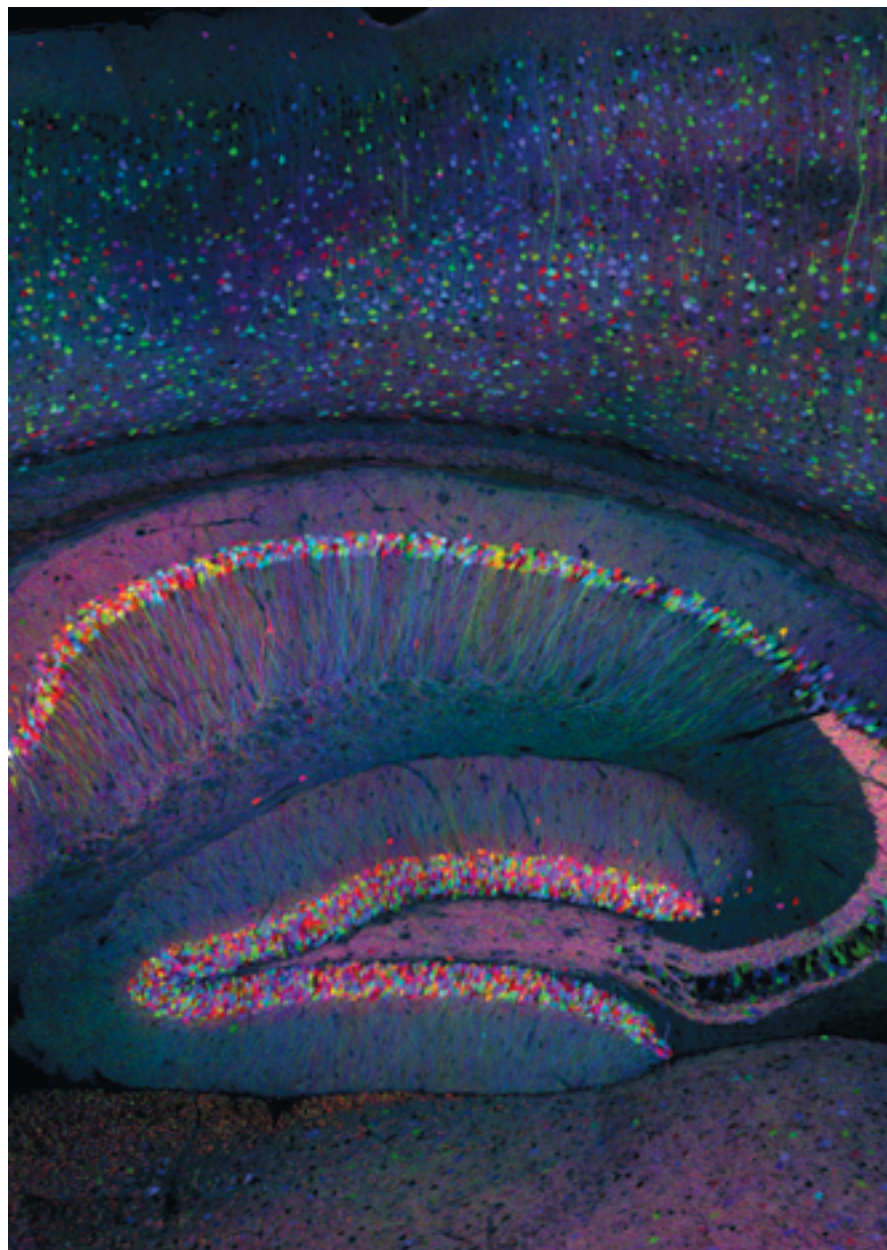
A slice through the brain of a mouse lights up in "brainbow" colors labeling various populations of neurons. The colors are generated by combinations of fluorescent proteins derived from genes introduced into the mouse germ line. The genes randomly recombine in a manner much like that of antibodies in the immune system, producing a wide spectrum of colors. A colored neuron looks like a balloon on a string, the balloon itself being the cell body and the string the long axonal fiber. The color contrast allows individual axons to be traced from source to destination. The image shows part of the cerebral cortex above and the hippocampus below, including the dentate gyrus at the bottom. Image courtesy of Jeff W. Lichtman of Harvard University.

average rate of spiking. The extent to which biological neurons adopt these simplifying strategies remains a matter of controversy and conjecture.

Remembering When

Still another challenge awaits when trying to interpret brain architecture through the lens of computer technology. Digital computers rely on directly addressed data storage. A pattern of bits is written to a specific location; subsequent reading of the same location retrieves the data. Neuroscientists have searched everywhere in the nervous system for such an addressable read-write memory, without success.

The prevailing model of information storage in the brain is *associative* memory. Conventional computer memory works like a coat-check room. When you hand over your coat,

you get a numbered ticket; later, when you present that ticket, you receive whatever coat is on the hanger with the matching number. Associative memory is a coat-check room that issues no tickets. To retrieve your coat, you list some of its attributes—blue wool jacket, missing top button—and the clerk brings out all the coats that match the description.

The brain's implementation of associative memory is thought to be based on the adjustment of synaptic transmission in response to experience, as proposed by Hebb. This mechanism seems well suited to storing perceptual and motor patterns that we learn by repeated exposure or rehearsal: memories of places and faces, the habitual motions of the fingers when tying shoelaces or playing guitar chords. Each repetition is believed to strengthen the synaptic connections between neurons that fire simultaneously. The eventual result is a "cell assembly," a set of neurons that tend to respond as a group whenever a sufficiently large subset of the assembly is stimulated.

Experiments are starting to reveal the biochemical nature of learning-induced changes in synapses. A harder question is how a shifting pattern of synaptic weights encodes an abstract concept or a narrative. Somewhere in my brain is an enduring record of the important fact that $7 \times 9 = 63$. There's also a memory of the long, dreary struggle to implant that fact—the hours spent drilling with flash cards. I am intensely curious about how those two kinds of knowledge are represented in my head.

What about memories that are formed without the need for practice or repetition? People can narrate the plot of a movie (often in excruciating detail) after a single viewing. For that matter, honeybees can remember and report the location of flowers after a single foraging trip. Can these feats also be explained by some variant of Hebbian learning?

The Small World of the Brain

The late Valentino Braitenberg took a distinctively quantitative approach to the great puzzles of neuroscience. With his colleagues at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, he spent more than two decades counting and measuring the cells of the mammalian cerebral cortex. The facts and figures he gathered did not solve all the mysteries, but any proposed answers will have to take his data into account.

A neuron in the human cortex (specifically, a pyramidal cell) has a frizzy nimbus of dendrites that extends throughout a volume of roughly 1 cubic millimeter. But the cell does not fill this entire volume; on the contrary, it shares the space with 100,000 other cells in a densely woven thatch of dendrites and axons. The combined length of all the dendrites in this volume is about 450 meters; the total length of the axons is even greater, more than 4 kilometers. In this tangled skein of nerve fibers, you might imagine that every cell would cross wires with every other cell. But Braitenberg found that most of the cells make no contact with one another. Choosing any two neurons that have fibers extending into the same cubic millimeter, the probability that they share a synapse is only 2 percent.

However, this sparse connectivity does not mean that the cortex breaks down into isolated clusters of cells. A signal from any neuron can reach all the other neurons in the entire cortex—all 20 billion of them—after passing through no more than two or three synapses. The cortex is a "small world" network, like social networks in which everyone is connected by chains of friends of friends.

On this evidence Braitenberg argued that the cortex is "a device for the most widespread diffusion and most intimate mixing of signals." He suggested that this architecture is just what would be expected in an associative memory. Incoming signals spread rapidly throughout the cortex, reaching nearly all the neurons. Each possible combination of inputs evokes a different response: One set of cells is activated when we encounter quacking and waddling creatures, another set recognizes the barking and tail-wagging ones.

Questions remain about how such a system would reliably distinguish a boundless spectrum of possible patterns. At one extreme, every concept is associated with a single neuron; this is the notion of the "grandmother cell," which lights up when granny enters the room. Cell assemblies might be seen as a broader version of the same idea, with overlapping populations of cells representing percepts and concepts. Braitenberg favored an even more diffuse scheme, in which concepts are embodied in the global state of the entire network.

Not everyone goes along with this view of the cortex as a large, undifferentiated memory organ. Indeed, there is much compelling evidence that regions have special functions, such as vision and speech. Braitenberg believed these opposing theories of the cortex could be reconciled.

Dream the Impossible Dream

Frances Crick, who took up neuroscience after conquering molecular biology, wrote in 1979 about the prospects for understanding the brain: "It is no use asking for the impossible, such as, say, the exact wiring diagram for a cubic millimeter of brain tissue and the way all its neurons are firing." Current brain-mapping projects do exactly what Crick believed impossible.

A goal of the Connectome Project at Harvard is to image the microanatomy of a cubic millimeter of mouse brain in sufficient detail to resolve individual synapses and create a full connectivity map. The first step is to slice the tiny block of tissue into 20,000 sections, each 50 nanometers thick. Each slice is imaged by an electron microscope with a resolution of 5 nanometers. The resulting data set will be about 800 terabytes. The second phase of the project is more difficult: Identifying features of cells in the individual images and correctly aligning the features in successive slices to reconstruct the full three-dimensional geometry. For these tasks to be completed at reasonable speed and cost, both phases of the operation must be automated. A group led by Jeff W. Lichtman and Hanspeter Pfister of Harvard has recently reported on a pilot project with a cube of tissue 30 micrometers on a side.

As for the second part of Crick's impossible request, a manifesto for the Brain Activity Map declares: "We propose to record every action potential from every neuron within a circuit—a task we believe is feasible." Admittedly, the task is not feasible with present instruments, which either average the activity of large ensembles of cells or isolate small numbers of single cells. One approach to bridging the gap would rely on arrays of nanoelectrodes to record simultaneously from many cells. The alternatives are optical techniques, in which molecules or nanoparticles implanted in neurons

emit light in response to ion flows or voltage changes.

The new European project, led by Henry Markram of the Swiss Federal Institute of Technology in Lausanne, seems to go well beyond Crick's impossible dream. Within a decade, Markram proposes to build a computer simulation of the entire human brain at a level of detail fine enough to include structural and physiological features of individual cells. Then he envisions linking the model brain to a virtual body in a virtual environment. He even mentions looking for cognitive abilities like those of a human infant.

Such a model can be built without first having a full wiring diagram of the brain, Markram says; its assembly will be guided by the same genetic and developmental rules that operate in the embryo. And it can be built without first ascertaining the nature of memory or the neural encoding of information; as a matter of fact, the simulation should help resolve those enigmas, according to Markram. Europe is now making a €1 billion bet that these grandiose plans will succeed.

Markram's Human Brain Project is not the first program with that name; the U.S. National Academy of Sciences launched an identically named research effort more than 20 years ago, when President George H. W. Bush declared that the 1990s would be "the decade of the brain." It's a little discouraging to be starting down the same path again, with the big questions still unanswered. But, as the owner of a brain that's still curious about itself, I believe the quest must go on.

Bibliography

Alivisatos, A., et al. 2012. The Brain Activity Map project and the challenge of functional connectomics. *Neuron* 74:970–974.

Braitenberg, V., and A. Schüz. 1998. *Cortex: Statistics and Geometry of Neuronal Connectivity*. Berlin: Springer.

Crick, F. H. C. 1979. Thinking about the brain. *Scientific American* 241(3):219–232.

Kaynig, V., et al. 2013 (preprint). Large-scale automatic reconstruction of neuronal processes from electron microscopy images. http://arxiv.org/abs/1303.7186.

Lichtman, J. W., and W. Denk. 2011. The big and the small: Challenges of imaging the brain's circuits. *Science* 334:618–623.

Markram, H., et al. 2012. *The Human Brain Project: A Report to the European Commission*. Lausanne, Switzerland: HBP-PS Consortium.

McCulloch, W. S., and W. H. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5:115–133.