Also by John Brockman

As Author:
By the Late John Brockman
37
Afterwords
The Third Culture: Beyond the Scientific Revolution
Digerati

As Editor:
About Bateson
Speculations
Doing Science
Ways of Knowing
Creativity
The Greatest Inventions of the Past 2,000 Years
The Next Fifty Years
The New Humanists
Curious Minds
What We Believe but Cannot Prove
My Einstein
Intelligent Thought
What Is Your Dangerous Idea?
What Are You Optimistic About?
What Have You Changed Your Mind About?
This Will Change Everything
Is the Internet Changing the Way You Think?
Culture
The Mind
This Will Make You Smarter
This Explains Everything
Thinking
What Should We Be Worried About?
The Universe

As Coeditor:
How Things Are (with Katinka Matson)

This Idea Must Die

Scientific Theories That Are Blocking Progress

Edited by John Brockman

HARPER PERENNIAL

NEW YORK ◦ LONDON ◦ TORONTO ◦ SYDNEY ◦ NEW DELHI ◦ AUCKLAND

GRANDMOTHER CELLS

TERRENCE J. SEJNOWSKI
Computational neuroscientist; Francis Crick Professor, Salk Institute for Biological Studies; coauthor (with Patricia S. Churchland),
The Computational Brain

In 2004, an epilepsy patient at the UCLA Medical Center whose brain was being monitored to detect the origin of the seizures was shown a series of pictures of celebrities. Electrodes implanted in the memory centers of the patient's brain reported spikes in response to the photos. One of the neurons responded vigorously to several pictures of Jennifer Aniston but not to other famous people. A neuron in another patient would respond only to pictures of Halle Berry, and even to her name, but not to pictures of Bill Clinton or Julia Roberts or the names of other famous people.

Such cells had been predicted fifty years ago, when it first became possible to record from single neurons in the brains of cats and monkeys. It was thought that in the hierarchy of visual areas of the cerebral cortex, the response properties of the neurons became more and more specific the higher the neuron was in the hierarchy—perhaps so specific that a single neuron would respond only to pictures of a single person. This became known as the grandmother-cell hypothesis, after the putative neuron in your brain that recognizes your grandmother. The team at UCLA seemed to have found such cells. Single neurons were also found that recognized specific objects and buildings, like the Sydney Opera House.

Despite this striking evidence, the grandmother-cell hypothesis is unlikely to be correct, or even a good explanation for these recordings. We're beginning to collect recordings from hundreds of cells simultaneously in mice, monkeys, and humans, and these are leading to a different theory for how the cortex perceives and decides. Nonetheless, the grandmother-cell hypothesis continues to have adherents, and the thinking that derives from focusing on single neurons still permeates the field of cortical electrophysiology. We'd make progress more quickly if we could retire the proverbial grandmother cell.

According to the grandmother-cell hypothesis, you perceive your grandmother when the cell is active, so it shouldn't fire to any other stimulus. Only a few hundred pictures were tested, and many more pictures were not tested, so we really don't know how selective the Jennifer Aniston cell was. Second, the likelihood that the electrode by chance happened to record from the only Jennifer Aniston neuron in the brain is low; it's more likely that there are many thousands. The same for the Halle Berry neuron, for everyone you know and every object you can recognize. There are many neurons in the brain, but not enough for each object and name that you know. An even deeper reason to be skeptical of the grandmother-cell hypothesis is that the function of a sensory neuron is only partially determined by its response to sensory inputs. Equally important is the output of the neuron and its effects downstream on behavior.

In monkeys, where it has been possible to record from many neurons simultaneously, stimuli and task-dependent signals are broadly distributed over large populations of neurons, each tuned to a different combination of features of the stimuli and task detail. The properties of such distributed representations were first studied in artificial neural networks in the 1980s.

Populations of simple model neurons called "hidden units" were trained to perform a mapping between a set of input units and a set of output units; these hidden units developed patterns of activity for each input that was highly distributed, like what has been observed in populations of cortical neurons. For example, the input units could represent faces from many different angles, and the output units could represent the names of the people. After being trained on many examples, each of the hidden units coded different combinations of features of the input units, such as fragments of eyes, noses, or head shapes.

A distributed representation can be used to recognize many versions of the same object, and the same set of neurons can recognize many different objects by differentially weighting their outputs. Moreover, the network can generalize by correctly classifying new inputs from outside the training set. Much more powerful versions of these early neural-network models, with more than twelve layers of hidden units in a hierarchy like that in our visual cortex, and using deep learning to adjust billions of synaptic weights, are now able to recognize tens of thousands of objects in images. This is a breakthrough in artificial intelligence, because performance continues to improve as the size of the network and the number of training examples increases. Companies worldwide are racing to build special-purpose hardware that would scale up these architectures. There's still a long way to go before the current systems approach the capacity of the human brain, which has a billion synapses in every cubic millimeter of cortex.

How many neurons are needed in a population that can discriminate between many similar objects, such as faces? From imaging studies, we know that many areas of the brain respond to faces, some with a high degree of selectivity. We'll need to

sample many neurons widely from these areas. The answer to this question may be a surprise, because there are also sound theoretical arguments for minimizing the numbers of neurons in the representation of an object. First, sparse coding would be more energy efficient. Second, learning a new object in the same population of neurons leads to interference with the others being represented in the population. An effective and efficient representation would be sparsely distributed.

In ten years, 1,000 times more neurons will be recorded and manipulated than is now possible, and new techniques are being developed to analyze them, which could lead to a deeper understanding of how activity in populations of neurons gives rise to thoughts, emotions, plans, and decisions. We may soon know the answer to the question of how many neurons represent an object or a concept in our brain—but will this retire the grandmother-cell hypothesis?