
Monopoly
tree
M&M's
tea
Legos
fireflies
Eiffel Tower
dog
Mother
Darth Vader
Buenos Aires
sister
Grandma
tire swing
Big Ben
Maradona
surfboard
Jennifer Aniston
Elm Street
Atlanta airport
mint julep

Rodrigo Quian Quiroga, a native of Argentina, is professor and head of the Bioengineering Research Group at the University of Leicester in England. He is author of the recently published *Borges and Memory: Encounters with the Human Brain* (MIT Press, 2012).

Itzhak Fried is a professor of neurosurgery and director of the Epilepsy Surgery Program at the U.C.L.A. David Geffen School of Medicine. He is also a professor at the Tel Aviv Sourasky Medical Center and Tel Aviv University.

Christof Koch is professor of cognitive and behavioral biology at the California Institute of Technology and chief scientific officer at the Allen Institute for Brain Science in Seattle.

NEUROSCIENCE

Brain Cells for Grandmother

Each concept—each person or thing in our everyday experience—may have a set of corresponding neurons assigned to it

By Rodrigo Quian Quiroga, Itzhak Fried and Christof Koch

IN BRIEF

For decades neuroscientists have debated how memories are stored. That debate continues today, with competing theories—one of which suggests that single neurons hold the recollection, say, of your grandmother or of a famous movie star.

The alternative theory asserts that each memory is distributed across many millions of neurons. A number of recent experiments during brain surgeries provide evidence that relatively small sets of neurons in specific regions are involved with the encoding of memories.

At the same time, these small groupings of cells may represent many instances of one thing; a visual image of Grandma's face or her entire body—even a front and side view or the voice of a Hollywood star such as Jennifer Aniston.

Photograph by Dan Saelinger

ONCE A BRILLIANT RUSSIAN NEUROSURGEON NAMED Akakhi Akakhievitch had a patient who wanted to forget his overbearing, impossible mother.

Eager to oblige, Akakhievitch opened up the patient's brain and, one by one, ablated several thousand neurons, each of which related to the concept of his mother. When the patient woke up from anesthesia, he had lost all notion of his mother. All memories of her, good and bad, were gone. Jubilant with his success, Akakhievitch turned his attention to the next endeavor—the search for cells linked to the memory of "grandmother."

The story, of course, is fiction. The late neuroscientist Jerry Lettvin (who, unlike Akakhievitch, was real) told it to a crowd of students at the Massachusetts Institute of Technology in 1969 to illustrate the provocative idea that as few as about 18,000 neurons could form the basis of any particular conscious experience, thought or memory of a relative or any other person or object we might come across. Lettvin never proved or disproved his audacious hypothesis, and for more than 40 years scientists have debated, mostly in jest, the idea of "grandmother cells."

The idea of neurons that store memories in such a highly specific manner goes all the way back to William James, who in the late 19th century conceived of "pontificial cells" to which our consciousness is attached. The existence of these cells, though, runs counter to the dominant view that the perception of any specific individual or object is accomplished by the collective activity of many millions if not billions of nerve cells, what Nobel laureate Charles Sherrington in 1940 called "a millionfold democracy." In this case, the activity of any one individual nerve cell is meaningless. Only the collaboration of very large populations of neurons creates meaning.

Neuroscientists continue to argue about whether it takes relatively few neurons—on the order of thousands or less—to serve as repositories for a particular concept or whether it takes hundreds of millions distributed widely throughout the brain. Attempts to resolve this dispute are leading to new understanding of the workings of memory and conscious thought—with a little help from Hollywood.

JENNIFER ANISTON NEURONS

SOME YEARS AGO—together with Gabriel Kreiman, now a faculty member at Harvard Medical School, and Leila Reddy, now a researcher at the Brain and Cognition Research Center in Toulouse, France—we performed experiments that led to the discovery of a neuron in the hippocampus of one patient, a brain region known to be involved in memory processes, that responded very strongly to different photographs of actress Jennifer Aniston but not to dozens of other actors, celebrities, places and animals. In another patient, a neuron in the hippocampus lit up at the sight of pictures of actress Halle Berry and even to her name written on the computer screen but responded to nothing else. Another neuron fired selectively to pictures of Oprah Winfrey and to her name written on the screen and spoken by a computer-synthesized voice. Yet another fired to pictures of Luke Skywalker and to his written and spoken name, and so on.

This kind of observation is made possible by the direct recording of the activity of individual neurons. Other more common techniques, such as functional brain imaging, can pinpoint activity throughout the brain when a volunteer performs a given task. Yet although functional imaging can track the overall power consumption of typically a few million cells, it cannot identify small groups of neurons, let alone individual cells. To record the electrical pulses emitted by individual neurons, microelectrodes thinner than a human hair need to be implanted in the brain. This technique is used less commonly than functional imaging, and only special medical circumstances warrant implantation of these electrodes in humans.

One of those rare circumstances occurs when treating patients with epilepsy. When seizures cannot be controlled with medication, these patients may be candidates for remedial surgery. The medical team examines clinical evidence that can pinpoint the location of the area where seizures start, the epileptic focus, which can potentially be surgically removed to cure the patient. Initially this evaluation involves noninvasive procedures, such as brain imaging, consideration of clinical evidence and the study of pathological electrical activity—a multitude of epileptic discharges that all occur in lockstep—with EEG recordings made from the patient's scalp. But when it is not possible to accurately determine the location of the epileptic focus with these methods, neurosurgeons may implant electrodes deep inside the skull to continuously monitor in the hospital brain activity over several days and then analyze the seizures observed.

Scientists sometimes ask patients to volunteer for research studies during the monitoring period, studies in which a variety of cognitive tasks are performed as brain activity is recorded. At the University of California, Los Angeles, we have employed a unique technique to record within the skull using flexible electrodes with tiny microwires; the technology was developed by one of us (Fried), who heads the Epilepsy Surgery Program at U.C.L.A. and collaborates with other scientists from

PRECEDING PAGES: DOMINIQUE BAYNES (*prop styling*)

around the world, including Koch's group at the California Institute of Technology and Quian Quiroga's laboratory at the University of Leicester in England. This technique furnishes an extraordinary opportunity to record directly from single neurons for days at a time in awake patients and provides the ability to study the firing of neurons during various tasks—monitoring the incessant chattering that occurs while patients look at images on a laptop, recall memories or perform other tasks. That is how we discovered the Jennifer Aniston neurons and unwittingly revived the debate ignited by Lettvin's parable.

GRANDMOTHER CELLS REVISITED

ARE NERVE CELLS such as the Jennifer Aniston neuron the long-debated grandmother cells? To answer that question, we have to be more precise about what we mean by grandmother cells. One extreme way of thinking about the grandmother cell hypothesis is that only one neuron responds to one concept. But if we could find one single neuron that fired to Jennifer Aniston, it strongly suggests that there must be more—the chance of finding the one and only one among billions is minuscule. Moreover, if only a single neuron would be responsible for a person's entire concept of Jennifer Aniston, and it were damaged or destroyed by disease or accident, all trace of Jennifer Aniston would disappear from memory, an extremely unlikely prospect.

A less extreme definition of grandmother cells postulates that many more than a solitary neuron respond to any one concept. This hypothesis is plausible but very difficult, if not impossible, to prove. We cannot try every possible concept to prove that the neuron fires only to Jennifer Aniston. In fact, the opposite is often the case: we often find neurons that respond to more than one concept. Thus, if a neuron fires only to one person during an experiment, we cannot rule out that it could have also fired to some other stimuli that we did not happen to show.

A single neuron that responded to Luke Skywalker and his written and spoken name also fired to the image of Yoda.

For example, the day after finding the Jennifer Aniston neuron we repeated the experiment, now using many more pictures related to her, and found that the neuron also fired to Lisa Kudrow, a costar in the TV series *Friends* that catapulted both to fame. The neuron that responded to Luke Skywalker also fired to Yoda, another Jedi from *Star Wars*; another neuron fired to two basketball players; another to one of the authors (Quian Quiroga) of this article and other colleagues who interacted with the patient at U.C.L.A., and so on. Even then, one can still argue that these neurons are grandmother cells that are responding to broader concepts, namely, the two blond women from *Friends,* the Jedis from *Star Wars,* the basketball players, or the scientists doing experiments with the patient. This expanded definition turns the discussion of whether these neurons should be considered grandmother cells into a semantic issue.

Let us leave semantics aside for now and focus instead on a few critical aspects of these so-called Jennifer Aniston neurons. First, we found that the responses of each cell are quite selective—each fires to a small fraction of the pictures of celebrities, politicians, relatives, landmarks, and so on, presented to the patient. Second, each cell responds to multiple representations of a particular individual or place, regardless of specific visual features of the picture used. Indeed, a cell fires in a similar manner in response to different pictures of the same person and even to his or her written or spoken name. It is as if the neuron in its firing patterns tells us: "I know it is Jennifer Aniston, and it does not matter how you present her to me, whether in a red dress, in profile, as a written name or even when you call her name out loud." The neuron, then, seems to respond to the concept—to any representation of the thing itself. Thus, these neurons may be more appropriately called concept cells instead of grandmother cells. Concept cells may sometimes fire to more than one concept, but if they do, these concepts tend to be closely related.

A CODE FOR CONCEPTS

TO UNDERSTAND the way a small number of cells become attached to a particular concept such as Jennifer Aniston, it helps to know something about the brain's complex processes for capturing and storing images of the myriad of objects and people encountered in the world around us. The information taken in by the eyes first goes—via the optic nerve leaving the eyeball—to the primary visual cortex at the back of the head. Neurons there fire in response to a tiny portion of the minute details that compose an image, as if each were lighting up like a pixel in a digital image or as if they were the colored dots in a pointillist painting by Georges Seurat.

One neuron does not suffice to tell whether the detail is part of a face, a cup of tea or the Eiffel Tower. Each cell forms part of an ensemble, a combination that generates a composite image presented, say, as *A Sunday Afternoon on the Island of La Grande Jatte.* If the picture changes slightly, some of the details will vary, and the firing of the corresponding set of neurons will change as well.

The brain needs to process sensory information so that it captures more than a photograph—it must recognize an object and integrate it with what is already known. From the primary visual cortex, the neuronal activation triggered by an image moves through a series of cortical regions toward more frontal areas. Individual neurons in these higher visual areas respond to entire faces or whole objects and not to local details. Just one of these high-level neurons can tell us that the image is a face and not the Eiffel Tower. If we slightly vary the picture, move it about or change the lighting illuminating it, it will change some features, but these neurons do not care much about small differences in detail, and their firing will remain more or less the same—a property known as visual invariance.

Neurons in high-level visual areas send their information to the medial temporal lobe—the hippocampus and surrounding cortex—which is involved in memory functions and is where we found the Jennifer Aniston neurons. The responses of neurons in the hippocampus are much more specific than in the higher visual cortex. Each of these neurons responds to a particular person or, more precisely, to the concept of that person: not only to the face and other facets of appearance but also to closely associated attributes such as the person's name.

In our research, we have tried to explore how many individual neurons fire to represent a given concept. We had to ask

whether it is just one, dozens, thousands or perhaps millions. In other words, how "sparse" is the representation of concepts? Clearly, we cannot measure this number directly, because we cannot record the activity of all neurons in a given area. Using statistical methods, Stephen Waydo, at the time a doctoral student with one of us (Koch) at Caltech, estimated that a particular concept triggers the firing of no more than a million or so neurons, out of about a billion in the medial temporal lobe. But because we use pictures of things that are very familiar to the patients in our research—which tend to trigger more responses—this number should be taken strictly as an upper bound; the number of cells representing a concept may be 10 or 100 times as small, perhaps close to Lettvin's guess of 18,000 neurons per concept.

Contrary to this argument, one reason to think that the brain does not code concepts sparsely, but rather distributes them across very large neuronal populations, is that we may not have enough neurons to represent all possible concepts and their variations. Do we, for instance, have a big enough store of brain cells to picture Grandma smiling, weaving, drinking tea or waiting at the bus stop, as well as the Queen of England greeting the crowds, Luke Skywalker as a child on Tatooine or fighting Darth Vader, and so on?

To answer this question, we should first consider that, in fact, a typical person remembers no more than 10,000 concepts. And this is not a lot in comparison to the billion nerve cells that make up the medial temporal lobe. Furthermore, we have good reason to think that concepts may be coded and stored very efficiently in a sparse way. Neurons in the medial temporal lobe just do not care about different instances of the same concept—they do not care if Luke is sitting or standing; they only care if a stimulus has something to do with Luke. They fire to the concept itself no matter how it is presented. Making the concept more abstract—firing to all instances of Luke—reduces the information that a neuron needs to encode and allows it to become highly selective, responding to Luke but not to Jennifer.

Simulation studies by Waydo underscore this view even further. Drawing on a detailed model of visual processing, Waydo built a software-based neural network that learned to recognize many unlabeled pictures of airplanes, cars, motorbikes and human faces. The software did so without supervision from a teacher. It was not told "this is a plane and that a car." It had to figure this out by itself, using the assumption that the immense variety of possible images is in reality based on a small number of people or things and that each is represented by a small subset of neurons, just as we found in the medial temporal lobe. By incorporating this sparse representation in the software simulation, the network learned to distinguish the same persons or objects even when shown in myriad different ways, a finding similar to our observations from human brain recordings.

CONCEPT CELLS

To Code a Memory

Neuroscientists ardently debate two alternative theories of how memories are encoded in the brain. One theory contends that the representation of a single memory—the image of Luke Skywalker, for instance—is stored as bits and pieces distributed across millions or perhaps billions of neurons. The alternative view, which has gained more scientific credibility in recent years, holds that a relatively few neurons, numbering in the thousands or perhaps even less, constitute a "sparse" representation of an image. Each of those neurons will switch on to the image of Luke, whether from a distance or close-up. Some but not all of the same group of neurons will also fire to the related image of Yoda. Similarly, a separate set of specific neurons activates when perceiving Jennifer Aniston.

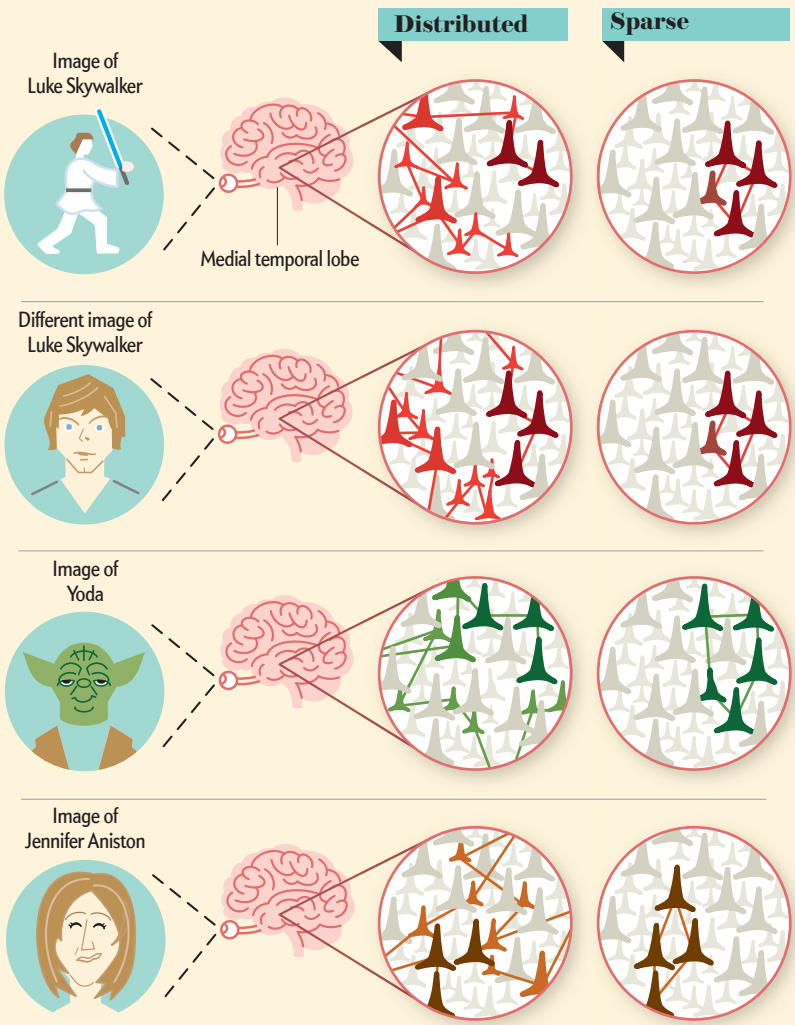

Illustration by Stephen Savage

WHY CONCEPT CELLS?

OUR RESEARCH is closely related to the question of how the brain interprets the outside world and translates perceptions into memories. Consider the famous 1953 case of patient H.M., who suffered from intractable epilepsy. As a desperate approach to try to stop his seizures, a neurosurgeon removed his hippocampus and adjoining regions in both sides of the brain. After the surgery, H.M. could still recognize people and objects and remember events that he had known before the surgery, but the unexpected result was that he could no longer make new long-lasting memories. Without the hippocampus, everything that happened to him quickly fell into oblivion. The 2000 movie *Memento* revolves around a character who has a similar neurological condition.

H.M's case demonstrates that the hippocampus, and the medial temporal lobe in general, is not necessary for perception but is critical for transferring short-term memories (things we remember for a short while) into long-term memories (things remembered for hours, days or years). In line with this evidence, we argue that concept cells, which reside in these areas, are critical for translating what is in our awareness—whatever is triggered by sensory inputs or internal recall—into long-term memories that will later be stored in other areas in the cerebral cortex. We believe that the Jennifer Aniston neuron we found was not necessary for the patient to recognize the actress or to remember who she was, but it was critical to bring Aniston into awareness for forging new links and memories related to her, such as later remembering seeing her picture.

Our brains may use a small number of concept cells to represent many instances of one thing as a unique concept—a sparse and invariant representation. The workings of concept cells go a long way toward explaining the way we remember: we recall Jennifer and Luke in all guises instead of remembering every pore on their faces. We neither need (nor want) to remember every detail of whatever happens to us.

What is important is to grasp the gist of particular situations involving persons and concepts that are relevant to us, rather than remembering an overwhelming myriad of meaningless details. If we run into somebody we know in a café, it is more important to remember a few salient events at this encounter than what exactly the person was wearing, every single word he used or what the other strangers relaxing in the café looked like. Concept cells tend to fire to personally relevant things because we typically remember events involving people and things that are familiar to us and we do not invest in making memories of things that have no particular relevance.

Memories are much more than single isolated concepts. A memory of Jennifer Aniston involves a series of events in which she—or her character in *Friends* for that matter—takes part. The full recollection of a single memory episode requires links between different but associated concepts: Jennifer Aniston linked to the concept of your sitting on a sofa while spooning ice cream and watching *Friends*.

If two concepts are related, some of the neurons encoding one concept may also fire to the other one. This hypothesis gives a physiological explanation for how neurons in the brain encode associations. The tendency for cells to fire to related concepts may indeed be the basis for the creation of episodic memories (such as the particular sequence of events during the café encounter) or the flow of consciousness, moving spontaneously from one concept to the other. We see Jennifer Aniston, and this perception evokes the memory of the TV, the sofa and ice cream—related concepts that underlie the memory of watching an episode of *Friends*. A similar process may also create the links between aspects of the same concept stored in different cortical areas, bringing together the smell, shape, color and texture of a rose—or Jennifer's appearance and voice.

Given the obvious advantages of storing high-level memories as abstract concepts, we can also ask why the representation of these concepts has to be sparsely distributed in the medial temporal lobe. One answer is provided by modeling studies, which have consistently shown that sparse representations are necessary for creating rapid associations.

The technical details are complex, but the general idea is quite simple. Imagine a distributed—as opposite of sparse—representation for the person we met in the café, with neurons coding for each minute feature of that person. Imagine another distributed representation for the café itself. Making a connection between the person and the café would require creating links among the different details representing each concept but without mixing them up with others, because the café looks like a comfortable bookstore and our friend looks like somebody else we know.

Creating such links with distributed networks is very slow and leads to the mixing of memories. Establishing such connections with sparse networks is, in contrast, fast and easy. It just requires creating a few links between the groups of cells representing each concept, by getting a few neurons to start firing to both concepts. Another advantage of a sparse representation is that something new can be added without profoundly affecting everything else in the network. This separation is much more difficult to achieve with distributed networks, where adding a new concept shifts boundaries for the entire network.

Concept cells link perception to memory; they give an abstract and sparse representation of semantic knowledge—the people, places, objects, all the meaningful concepts that make up our individual worlds. They constitute the building blocks for the memories of facts and events of our lives. Their elegant coding scheme allows our minds to leave aside countless unimportant details and extract meaning that can be used to make new associations and memories. They encode what is critical to retain from our experiences.

Concept cells are not quite like the grandmother cells that Lettvin envisioned, but they may be an important physical basis of human cognitive abilities, the hardware components of thought and memory. SA

MORE TO EXPLORE

Sparse but Not "Grandmother-Cell" Coding in the Medial Temporal Lobe. R. Quian Quiroga, G. Kreiman, C. Koch and I. Fried in *Trends in Cognitive Sciences,* Vol. 12, No. 3, pages 87–91; March 2008.

Percepts to Recollections: Insights from Single Neuron Recordings in the Human Brain. Nanthia Suthana and Itzhak Fried in *Trends in Cognitive Sciences,* Vol. 16, No. 8, pages 427–436; July 16, 2012.

Concept Cells: The Building Blocks of Declarative Memory Functions. Rodrigo Quian Quiroga in *Nature Reviews Neuroscience,* Vol. 13, pages 587–597; August 2012.

SCIENTIFIC AMERICAN ONLINE

Read an excerpt of Quian Quiroga's book on memory at ScientificAmerican.com/feb2013/brain-cells