Being John Malkovich

An advanced brain-machine interface enables patients to control individual nerve cells deep inside their own brain

BY CHRISTOF KOCH

"Cerebroscope" that shows what people are thinking is fictional—but science is finding ways to track thoughts.

IN PHILOSOPHY OF MIND, a "cerebroscope" is a fictitious device, a brain-computer interface in today's language, which reads out the content of somebody's brain. An autocerebroscope is a device applied to one's own brain. You would be able to see your own brain in action, observing the fleeting bioelectric activity of all its nerve cells and thus of your own conscious mind. There is a strange loopiness about this idea. The mind observing its own brain gives rise to the very mind observing this brain. How will this weirdness affect the brain? Neuroscience has answered this question more quickly than many thought possible. But first, a bit of background.

Epileptic seizures—hypersynchronized, self-maintained neural discharges that can sometimes engulf the entire brain—are a common neurological disorder. These recurring and episodic brain spasms are kept in check with drugs that dampen excitation and boost inhibition in the underlying circuits. Medication does not always work, however. When a localized abnormality, such as scar tissue or developmental miswiring, is suspected of triggering the seizure, neurosurgeons may remove the offending tissue.

To minimize side effects, it is vital to pinpoint the location from which the seizures originate; neuropsychological testing, brain scans and EEGs aid this determination. But if no structural pathologies are apparent from the outside, doctors begin with an invasive procedure. The neurosurgeon inserts a dozen or so electrodes into the soft tissue of the brain, via small holes drilled through the skull, and leaves them in place for a week or so. During this time, the patient lives and sleeps in the hospital ward, and the signals from the wires are monitored continuously. When a seizure occurs, doctors triangulate the origin of the aberrant electrical activity. Subsequent destruction or removal of the offending chunk of tissue reduces the number of seizures—sometimes eliminating them entirely.

Neurosurgeon and neuroscientist Itzhak Fried of the David Geffen School of Medicine at U.C.L.A. is one of the world's foremost specialists in this demanding trade, which requires great technical finesse. Fried and his colleagues perfected a variant of epilepsy monitoring in which the electrodes are hollowed out. This alteration permits them to insert tiny wires straight into the gray matter. Using appropriate electronics and fancy signal-detection algorithms, these miniaturized electrodes pick up the faint chattering of a bevy of just 10 to 50 neurons from the ceaseless background cacophony of the electrical activity of billions of cells.

From Senses to Memories

Under Fried's supervision, a group from my laboratory—Rodrigo Quian Quiroga, Gabriel Kreiman and Leila Reddy—discovered a remarkable set of neurons in the jungles of the medial temporal lobe, the source of many epileptic seizures. This region, deep inside the brain, which includes the hippocampus, turns visual and other sensory percepts into memories.

We enlisted the help of several epileptic patients. While they waited for their seizures, we showed them about 100 pictures of familiar people, animals, landmark buildings and objects. We hoped one or more of the photographs would prompt some of the monitored neurons to fire a burst of action potentials. Most of the time the search turned up empty-handed, although sometimes we would come upon neurons that responded to categories of objects, such as animals, outdoor scenes or faces in general. But a few neurons were much more discerning. One hippocampal neuron responded only

CHRISTOF KOCH (*Koch*); MORAN CERF (*Monroe projection*) to photos of actress Jennifer Aniston but not to pictures of other blonde women or actresses; moreover, the cell fired in response to seven very different pictures of Jennifer Aniston. We found cells that responded to images of Mother Teresa, to cute little animals and to the Pythagorean theorem, $a^2 + b^2 = c^2$.

Such cells, together with their sisters—for there are probably thousands of such cells in the medial temporal lobe for any one idea—encode a concept, such as Jennifer Aniston, no matter whether the patient sees or hears her name or looks at her picture. Think of them as the cellular substrate of the Platonic ideal of Jennifer Aniston. Whether the actress is sitting or running, whether her hair is up or down, as long as the patient recognizes Jennifer Aniston, those neurons are active.

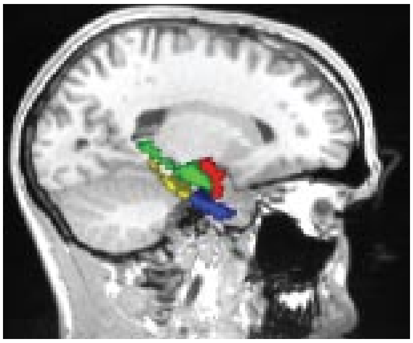
Nobody is born with cells selective for Jennifer Aniston. Like a sculptor patiently releasing a *Venus de Milo* or *Pietà* out of blocks of marble, the learning algorithms of the brain sculpt the synaptic fields in which concept neurons are embedded. Every time you encounter a particular person or object, a similar pattern of spiking neurons is generated in higher-order cortical regions. The networks in the medial temporal lobe recognize such repeating patterns and dedicate specific neurons to them. You have concept neurons that encode family members, pets, friends, coworkers, the politicians you watch on TV, your laptop, that painting you adore.

Conversely, you do not have concept cells for things you rarely encounter, such as the barista who just handed you a nonfat chai latte tea. If you were to befriend her, meet her later in a bar and let her into your life, the networks in the medial temporal lobe would recognize that the same pattern of spikes occurred repeatedly and would wire up concept cells to represent her.

Concept cells demonstrate compellingly that the specificity of conscious experience has a direct counterpart at the cellular level. Say you are recalling the iconic scene of Marilyn Monroe standing on a subway grill, trying to keep the wind from blowing her skirt up. This conscious percept will be caused by a coalition of neurons numbering perhaps in the hundreds or thousands rather than in the billions, as is commonly assumed.

Making Concepts Visible

More recently, Moran Cerf and others from my lab, together with Fried, hooked several concept cells to an external display to visualize a patient's thoughts. The idea is deceptively simple but fiendishly difficult to implement. It required three years of effort by Cerf, a computer-security specialist and a moviemaker turned Caltech graduate student, to pull off this feat. Let me walk you through one example. Cerf recorded from a neuron that fired in response to images of actor Josh Brolin (whom the patient knew from her favorite movie, *The Goonies*) and from another neuron that fired in response to the Marilyn Monroe scene I just mentioned. The patient looked at a monitor where these two images were superimposed, with the activity of the two cells controlling the extent to which she saw Brolin or Monroe in the hybrid image.



Four regions of the medial temporal lobe (*color highlights*) were sampled by a surgeon's electrodes.

Whenever the patient focused her thoughts on Brolin, the associated neuron fired more strongly. Cerf arranged the feedback such that the more this cell fired relative to the other one, the more visible Brolin became and the more the image of Monroe faded, and vice versa. The image on the screen kept changing until only Brolin or only Monroe remained visible and the trial was over. The patient loved it, as she felt that she controlled the movie purely with her thoughts. When she focused on Monroe, the associated neurons increased their firing rate, the cells for the competing concept, Brolin, dampened their activity, whereas the vast majority of neurons remained unaffected.

It might appear as if there are two people involved in this experiment, the way the puppeteer Craig occupied the head of actor John Malkovich in the 1999 movie *Being John Malkovich*. One is the patient's mind, instructing her brain to think of Monroe. The other is the one that is acting out the mind's desire—namely, the nerve cells in the medial temporal lobe that up- and down-regulate their activity accordingly. But both are part of the same brain. So who is in control of whom? Who is the puppeteer, and who the puppet?

All the weirdness of the mind-body nexus is apparent here. The patient doesn't feel an itch every time the Monroe neuron fires; she doesn't think, "Inhibition, inhibition, inhibition," to banish Brolin from the screen. She has absolutely no idea whatsoever what goes on inside her head. Yet the thought of Monroe translates into a particular pattern of neuronal activity. Events in her phenomenal mind find their parallel in her material brain. A mind-quake occurs simultaneously with a brain-quake. **M**

CHRISTOF KOCH is Lois and Victor Troendle Professor of Cognitive and Behavioral Biology at the California Institute of Technology. He serves on *Scientific American Mind*'s board of advisers.

(Further Reading)

- **Invariant Visual Representation by Single Neurons in the Human Brain.** R. Quian Quiroga, L. Reddy, G. Kreiman, C. Koch and I. Fried in *Nature,* Vol. 435, pages 1102–1107; June 23, 2005.
- **On-line, Voluntary Control of Human Temporal Lobe Neurons.** M. Cerf et al. in *Nature,* Vol. 467, pages 1104–1108; October 28, 2010.
- **Consciousness: Confessions of a Romantic Reductionist.** Christof Koch. Roberts Publishers, 2011.

JULIEN DUBOIS