

Brain regions that process faces reveal deep insights into the neural mechanisms of vision

By Doris Y. Tsao

WHEN I WAS IN HIGH SCHOOL, I LEARNED ONE DAY ABOUT THE DENSITY OF CURVES in an introductory course on calculus. A simple pair of differential equations, which model the interactions of predators and prey, can give rise to an infinite number of closed curves—picture concentric circles, one nested within another, like a bull’s-eye. What is more, the density of these curves varies depending on their location.

This last fact seemed so strange to me. I could easily imagine a finite set of curves coming close together or pulling apart. But how could an infinity of curves be denser in one region and less dense in another? I soon learned that there are different types of infinity, which have paradoxical qualities, like Hilbert’s Hotel (where the rooms are always fully booked but new guests can always be accommodated) and the Banach-Tarski apple (which can be split into five pieces and rearranged to make two apples of equal volume as the original). I spent hours poring over these mathematical proofs. Ultimately they struck me as symbolic magic of no real consequence, but the seed of interest had taken root.



Doris Y. Tsao is a professor of biology at the California Institute of Technology and an investigator of the Howard Hughes Medical Institute. She is also director of the Tianqiao and Chrissy Chen Center for Systems Neuroscience at Caltech. In October she was named a MacArthur Fellow.

IN BRIEF

Understanding vision

remains one of the grand challenges that neuroscientists confront.

One key aspect of this problem relates to the way the brain identifies faces, the most important social emblem.

Neurons in defined sections of the cerebral cortex, called face patches, are dedicated to recognizing faces.

Uncovering the organization of the face-patch system served as a prelude to deducing the underlying computations that the brain makes to identify faces.

This neural code may serve as a Rosetta stone for representing other objects besides faces.

Later, as an undergraduate at the California Institute of Technology, I learned about the experiments of David Hubel and Torsten Wiesel and their landmark discovery of how a region in the brain called the primary visual cortex extracts edges from the images relayed from the eyes. I realized that what had actually mystified me back in high school was the act of trying to *imagine* different densities of infinity. Unlike the mathematical tricks I had studied in high school, the edges that Hubel and Wiesel described are processed by neurons, so they actually exist in the brain. I came to recognize that visual neuroscience was a way to understand how this neural activity gives rise to the conscious perception of a curve.

The sense of excitement this realization triggered is hard to describe. I believe at each stage in life one has a duty. And the duty of a college student is to dream, to find the thing that captures one's heart and seems worth devoting a whole life to. Indeed, this is the single most important step in science—to find the right problem. I was captivated by the challenge of understanding vision and embarked on a quest to learn how patterns of electrical activity in the brain are able to encode perceptions of visual objects—not just lines and curves but even objects as hard to define as faces. Accomplishing this objective required pinpointing the specific brain regions dedicated to facial recognition and deciphering their underlying neural code—the means by which a pattern of electrical impulses allows us to identify people around us.

The journey of discovery began in graduate school at Harvard University, where I studied stereopsis, the mechanism by which depth perception arises from differences between the images in the two eyes. One day I came across a paper by neuroscientist Nancy Kanwisher, now at the Massachusetts Institute of Technology, and her colleagues, reporting the discovery of an area in the human brain that responded much more strongly to pictures of

faces than to images of any other object when a person was inside a functional magnetic resonance imaging (fMRI) brain scanner. The paper seemed bizarre. I was used to the brain being made of parts with names like basal ganglia and orbitofrontal cortex that had some vague purpose one could only begin to fathom. The concept of an area specifically devoted to processing faces seemed all too comprehensible and therefore impossible. Anyone could make a reasonable conjecture about the function of a face area—it should probably represent all the different faces that we know and something about their expression and gender.

As a graduate student, I had used fMRI on monkeys to identify areas activated by the perception of three-dimensionality in images. I decided to show pictures of faces and other objects to a monkey. When I compared activation in the monkey's brain to faces with activation to other objects, I found several areas that lit up selectively to faces in the temporal lobe (the area underneath the temple)—specifically in a region called the inferotemporal (IT) cortex. Charles Gross, a pioneer in the field of object vision, had discovered face-selective neurons in the IT cortex of macaques in the early 1970s. But he had reported that these cells were randomly scattered throughout the IT cortex. Our fMRI results provided the first indication that face cells might be concentrated into defined regions.

FACE PATCHES

AFTER PUBLISHING MY WORK, I was invited to give a talk describing the fMRI study for a faculty position at Caltech, but I was not offered the job. Many people were skeptical of the value of fMRI, which measures local blood flow, the brain's plumbing. They argued that showing increased blood flow to a brain area when a subject is looking at faces falls far short of clarifying what neurons in the area are actually encoding because the relation between blood flow and electrical activity is unclear. Perhaps by chance these face patches simply contained

a slightly larger number of neurons responsive to faces, like icebergs randomly clustered at sea.

Because I had done the imaging experiment in a monkey, I could directly address this concern by inserting an electrode into an fMRI-identified face area and asking, What images drive single neurons in this region most strongly? I performed this experiment together with Winrich Freiwald, then a postdoctoral fellow in Margaret Livingstone's laboratory at Harvard, where I was a graduate student. We presented faces and other objects to a monkey while amplifying the electrical activity of individual neurons recorded by the electrode. To monitor responses in real time, the neurons' electrical signals were converted to an audio signal that we could hear with a loudspeaker in the lab.

This experiment revealed an astonishing result: almost every single cell in the area identified through fMRI was dedicated to processing faces. I can recall the excitement of our first recording, hearing the “pop” of cell after cell responding strongly to faces and very little to other objects. We sensed we were on to something important, a piece of cortex that could reveal the brain's high-level code for visual objects. Marge remarked on the face patches: “You've found a golden egg.”

I also remember feeling surprised during that first experiment. I had expected the face area would contain cells that responded selectively to specific individuals, analogous to orientation-selective cells in the primary visual cortex that each respond to a specific edge orientation. In fact, a number of well-publicized studies had suggested that single neurons can be remarkably selective for the faces of familiar people—responding, say, only to Jennifer Aniston. Contrary to my expectation, each cell seemed to fire vigorously to almost any face.

I plugged madly away at Photoshop during these early experiments and found that the cells responded not just to faces of humans and monkeys but even to highly simplified cartoon faces.

Observing this phenomenon, I decided to create cartoon faces with 19 different features that seemed pertinent to defining the identity of a face, including inter-eye distance, face aspect ratio and mouth height, among other characteristics. We then went on to alter these values—the inter-eye distance, for instance, varied from being almost cyclopean to just inside the face boundary. Individual cells responded to most faces but interestingly did not always exhibit the exact same rate of firing to all faces. Instead there was a systematic variation in their response: when we plotted the firing of cells for the different cartoon features, we found a pattern in which there was a minimal response to one feature extreme—the smallest inter-eye distance, for instance—and a maximal response to the opposite extreme—the largest eye separation—with intermediate responses to feature values in the middle. The response as a function of the value for each feature looked like a ramp, a line slanted up or down.

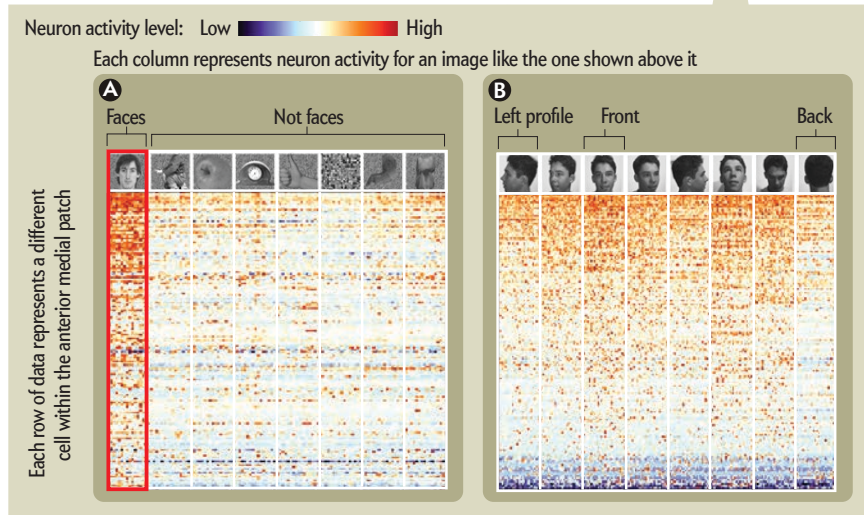
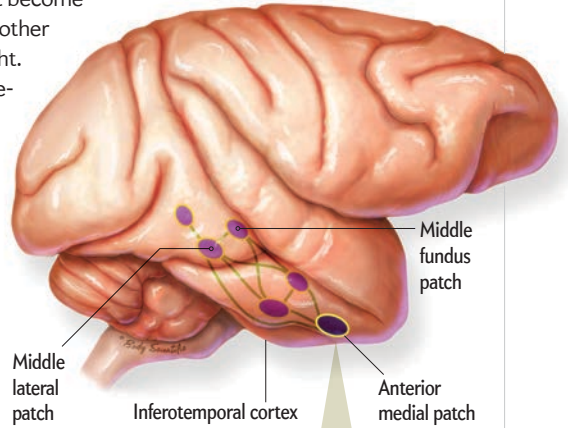
Once again, I was invited to give a job talk at Caltech. Returning, I had more to offer than just fMRI images. With the addition of the new results from single-cell recordings, it was clear to everyone that these face patches were real and likely played an important role in facial recognition. Furthermore, understanding their underlying neural processes seemed like an effective way to gain traction on the general problem of how the brain represents visual objects. This time I was offered the job.

CONTRAST IS KEY

AT CALTECH, my colleagues and I dug deeper into the question of how these cells detect faces. We took inspiration from a paper by Pawan Sinha, a vision and computational neuroscientist at M.I.T., that suggested faces could be discerned by checking for specific contrast relations between different regions of the face—whether the forehead region is brighter than the mouth region, for example. Sinha suggested a clever way to determine *which*

Where Are the Face Detectors?

A set of six nodes in the inferotemporal (IT) cortex of both brain hemispheres specializes in identifying faces. These “face patches” function as an assembly line: in the middle lateral and middle fundus patches, one neuron might become active when faces look straight ahead; another might turn on for a face looking to the right. At the end of the assembly line, in the anterior medial patch, varying views are stitched together. Neurons in this patch are active in response to the face of a specific individual, no matter if the view is from the front or side. Responses from a face patch of one monkey are generated for faces but not objects (red areas in **A**) and for the same individual, such as the dark-haired man, from varying angles (red areas in **B**).



contrast relations can be used to recognize a face: they should be the ones that are immune to changes in lighting. For example, left-eye-darker-than-nose is a useful feature for detecting a face because it does not matter if a face is photographed with lighting from above, left, right or below: the left eye is *always* darker than the nose (check for yourself).

From a theoretical standpoint, this idea provides a simple, elegant computational mechanism for facial recognition, and we wondered whether face cells might be using it. When we measured the response of cells to faces in which different regions varied in brightness, we found

that cells often had a significant preference for a particular contrast feature in an image.

To our astonishment, almost all the cells were wholly consistent in their contrast preferences—just a single cell was found that preferred the opposite polarity. Moreover, the preferred features were precisely those identified by Sinha as being invulnerable to lighting changes. The experiment thus confirmed that face cells use contrast relations to detect faces.

More broadly, the result confirmed that these cells truly were face cells. At talks, skeptics would ask, how do you know? You can't test every possible stimulus. How

can you be sure it's a face cell and not a pomegranate cell or a lawn mower cell? This result nailed it for me. The precise match between the way cells reacted to changes in contrast between different parts of the face and Sinha's computational prediction was uncanny.

Our initial experiments had revealed two nearby cortical patches that lit up to faces. But after further scanning (with the help of a contrast agent that increased several-fold the robustness of the signal), it became clear that there are actually six face patches in each of the brain's two hemispheres (making a dozen golden eggs total). They are distributed along the entire length of the temporal lobe. These six

patches, moreover, are not randomly scattered throughout the IT cortex. They are located in similar locations across hemispheres in each animal. Moreover, work by our group and others has found that a similar pattern of multiple face patches spanning the IT cortex exists in humans and other primates such as marmosets.

This observation of a stereotyped pattern suggested that the patches might constitute a kind of assembly line for processing faces. If so, one would expect the six patches to be connected to one another and each patch to serve a distinct function.

To explore the neural connections among patches, we electrically stimulated different patches with

tiny amounts of current—a technique called microstimulation—while the monkey was inside an fMRI scanner. The goal was to find out what other parts of the brain light up when a particular face patch is stimulated. We discovered that whenever we stimulated one face patch, the other patches would light up but not the surrounding cortex, indicating that, indeed, the face patches are strongly interconnected. Furthermore, we found that each patch performs a different function. We presented pictures of 25 people, each at eight different head orientations, to monkeys and recorded responses from cells in three regions: the middle lateral and middle fundus patches (ML/MF), the anterior lateral patch (AL) and the anterior medial patch (AM).

We found striking differences among these three regions. In ML/MF, cells responded selectively to specific views. For example, one cell might prefer faces looking straight ahead, whereas another might opt for faces looking to the left. In AL, cells were less view-specific. One class of cells responded to faces looking up, down and straight ahead; another responded to faces looking to the left or right. In AM, cells responded to specific individuals regardless of whether the view of the face was frontal or in profile. Thus, at the end of the network in AM, view-specific representations were successfully stitched into a view-invariant one.

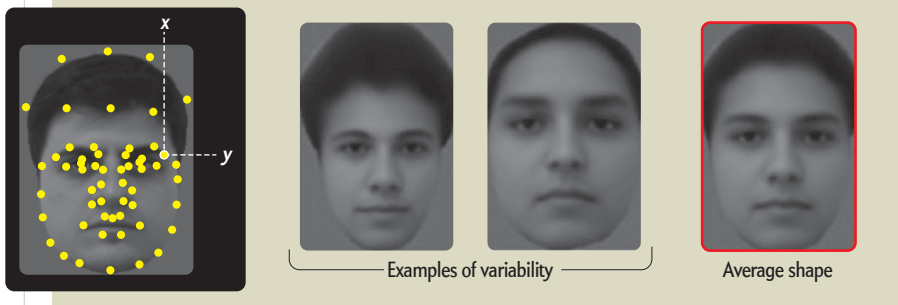
Apparently face patches do act as an assembly line to solve one of the big challenges of vision: how to recognize things around us despite changes in the way they look. A car can have any make and color, appear at any viewing angle and distance, and be partially obscured by closer objects such as trees or other cars. Recognizing an object despite these visual transformations is called the invariance problem, and it became clear to us that a major function of the face-patch network is to overcome this impediment.

Given the great sensitivity of cells in face patches to changes in facial identity, one might expect

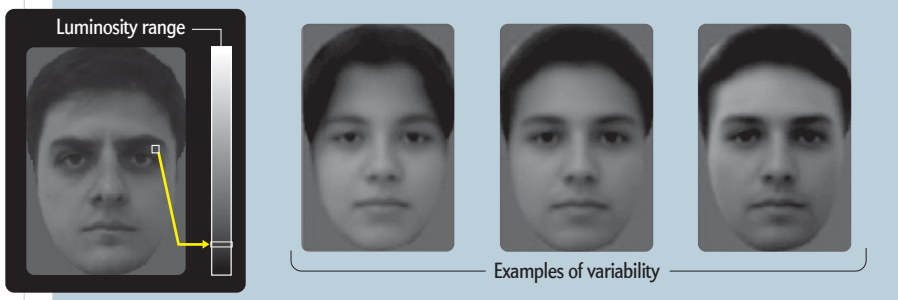
Shape + Appearance = Face

Identifying the face patches was only a first step. It then became necessary to explore what happens in the neurons within each patch, setting off a search for the brain's coding scheme for faces. To derive quantitative measures for faces, the Tsao laboratory came up with 25 features for shape and 25 for appearance that could be used by each neuron in a face patch—a 50-dimensional face space. The shape features can be thought of as those defining the skeleton—how wide the head is or the distance between the eyes. The appearance features specify the face's surface texture (complexion, eye or hair color, and so on).

Shape: Described by the position (x,y coordinates) of feature landmarks (yellow dots)



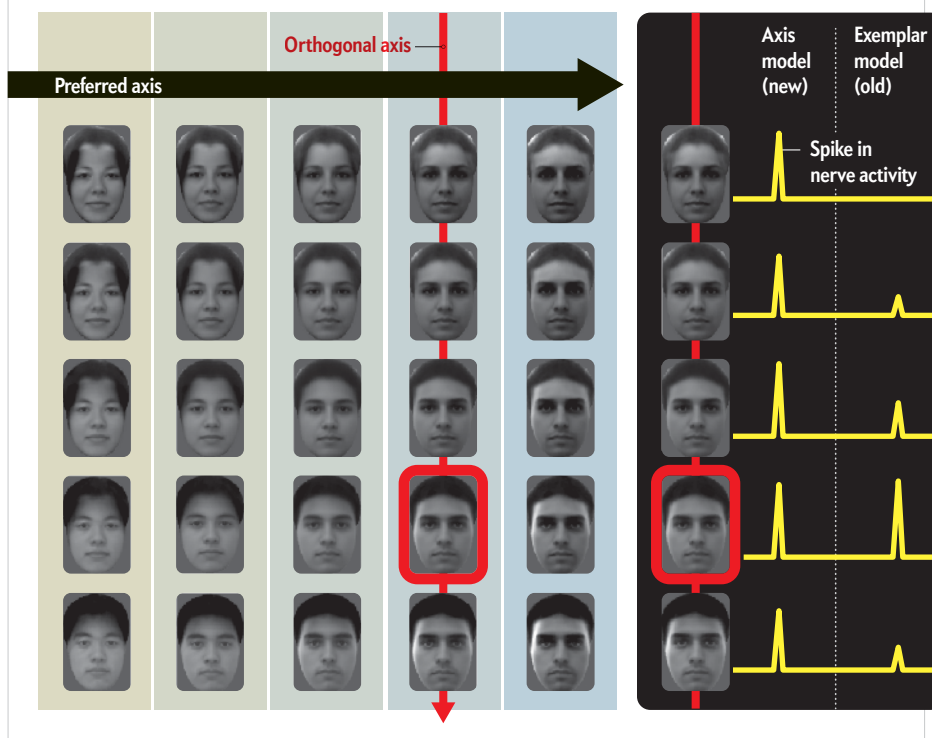
Appearance: Variations in luminosity of the image after first aligning it to match an average face shape



FROM "THE CODE FOR FACIAL IDENTITY IN THE PRIMATE BRAIN," BY LE CHANG AND DORIS Y. TSAO, IN CELL VOL. 169, NO. 6, JUNE 1, 2017 (face images)

The Face Code, at Last

Having 50 coordinates that describe shape and appearance allows for a description of neurons' firing in response to a particular face—a description that functions as a code that can be visualized geometrically. In this code, each face cell receives inputs for a face in the form of the 50 coordinates, or dimensions. The neuron then fires with a particular intensity in response to a certain face (red outlines), along what is called the preferred axis. The intensity increases steadily (monotonically) along the preferred axis. Furthermore the response is the same for every face on an axis at right angles to the preferred axis, even though those faces may look very different. This axis model of facial coding differs from a previous exemplar model that suggests that each neuron fires with maximum intensity to a single most preferred face.



that altering these cells' responses should modify an animal's perception of facial identity. Neuroscientists Josef Parvizi and Kalanit Grill-Spector of Stanford University had electrically stimulated a face-patch area in human subjects who had electrodes implanted in their brains for the purpose of identifying the source of epileptic seizures and found that stimulation distorted the subjects' perception of a face.

We wondered whether we would find the same effect in monkeys when we stimulated their face patches. Would doing so alter the perception only of faces, or would it affect that of other objects as well? The boundary between a face and a nonface object is fluid—one can see a face in a cloud or an electrical outlet if prompted. We wanted to use electrical microstimulation as a tool to delineate precisely what constitutes a face for a face patch. We trained monkeys to report whether two sequentially presented faces were the same or different. Consistent with the earlier results in humans, we found that microstimulation of face patches strongly distorted perception so that the animal would always signal two identical faces as being different.

Interestingly, microstimulation had no effect on the perception of many nonface objects, but it did significantly affect responses to a few objects whose shape is consistent with a face—apples, for one. But why does this stimulation influence the perception of an apple?

One possibility is that the face patches are typically used to represent not just faces but also other round objects like apples. Another hypothesis is that face patches are not normally used to represent these objects, but stimulation induces an apple to appear facelike. It remains unclear whether face patches are useful for detecting any nonface objects.

CRACKING THE CODE

UNCOVERING the organization of the face-patch system and properties of the cells within was a major accomplishment. But my dream when we

first began recording from face patches was to achieve something more. I had intuited that these cells would allow us to crack the neural code for facial identity. That means understanding how individual neurons process faces at a level of detail that would let us predict a cell's response to any given face or decode the identity of an arbitrary face based only on neural activity.

The central challenge was to figure out a way to describe faces quantitatively with high precision. Le Chang, then a postdoc in my lab, had the brilliant insight to adopt a technique from the field of computer vision called the active appearance model. In this approach, a face has two sets of descriptors, one for

shape and another for appearance. Think of the shape features as those defined by the skeleton—how wide the head is or the distance between the eyes. The appearance features define the surface texture of the face (complexion, eye or hair color, and so on).

To generate these shape and appearance descriptors for faces, we started with a large database of face images. For each face, we placed a set of markers on key features. The spatial locations of these markers described the shape of the face. From these varied shapes, we calculated an average face. We then morphed each face image in the database so its key features exactly matched that of the average face.

Pictures Worth 205 Neurons

For a given face, we can predict how a cell will respond by taking a weighted sum of all 50 face coordinates. To predict what face the monkey saw from neuronal activity, this entire process can be reversed: By knowing the response of 205 face cells, it is possible to predict the 50 coordinates defining the exact facial features—and make a highly accurate reconstruction of a given face.

Original Images from the Face Database



Corresponding Reconstructed Faces Based on Neuron Activity

The resulting images constituted the appearance of the faces independent of shape.

We then performed principal components analysis independently on the shape and appearance descriptors across the entire set of faces. This is a mathematical technique that finds the dimensions that vary the most in a complex data set.

By taking the top 25 principal components for shape and the top 25 for appearance, we created a 50-dimensional face space. This space is similar to our familiar 3-D space, but each point represents a face rather than a spatial location, and it comprises much more than just three dimensions. For 3-D space, any point can be described by three coordinates (x,y,z) . For a 50-D face space, any point can be described by 50 coordinates.

In our experiment, we randomly drew 2,000 faces and presented them to a monkey while recording cells from two face patches. We found that almost every cell showed graded responses—resembling a ramp slanting up or down—to a subset of the 50 features, consistent with my earlier experiments with cartoon faces. But we had a new insight about why this is important. If a face cell has ramp-shaped tuning to different features, its response can be roughly approximated by a simple weighted sum of the facial features, with weights determined by the slopes of the ramp-shaped tuning functions. In other words:

$$\text{response of face cells} = \text{weight matrix} \times 50 \text{ face features}$$

We can then simply invert this equation to convert it to a form that

lets us *predict* the face being shown from face cell responses:

$$50 \text{ face features} = (1/\text{weight matrix}) \times \text{response of face cells}$$

At first, this equation seemed impossibly simple to us. To test it, we used responses to all but one of the 2,000 faces to learn the weight matrix and then tried to predict the 50 face features of the excluded face. Astonishingly, the prediction turned out to be almost indistinguishable from the actual face.

A WIN-WIN BET

AT A MEETING in Ascona, Switzerland, I presented our findings on how we could reconstruct faces using neural activity. After my talk, Rodrigo Quian Quiroga, who discovered the famous Jennifer Aniston cell in the human medial temporal lobe in 2005 and is now at the University of Leicester in England, asked me how my cells related to his concept that single neurons react to the faces of specific people. The Jennifer Aniston cell, also known as a grandmother cell, is a putative type of neuron that switches on in response to the face of a recognizable person—a celebrity or a close relative.

I told Rodrigo I thought our cells could be the building blocks for his cells, without thinking very deeply about how this would work. That night, sleepless from jet lag, I recognized a major difference between our face cells and his. I had described in my talk how our face cells computed their response to weighted sums of different face features. In the middle of the night, I realized this computation is the same as a mathematical operation known as the dot product, whose geometric representation is the projection of a vector onto an axis (like the sun projecting the shadow of a flagpole onto the ground).

Remembering my high school linear algebra, I realized this implied that we should be able to construct a large “null space” of faces for each cell—a series of faces of varying identity that lie on an axis

perpendicular to the axis of projection. Moreover, all these faces would cause the cell to fire in exactly the same way.

And this, in turn, would suggest cells in face patches are fundamentally different from grandmother cells. It would demolish the vague intuition everyone shared about face cells—that they should be tuned to specific faces.

I was the first person in the meeting's breakfast hall at 5 A.M. the next morning and hoped to find Rodrigo so I could tell him about this counterintuitive prediction. Amazingly, when he finally showed up, he told me he had the exact same idea. So we made a bet, and Rodrigo allowed the terms to be framed in a way that would be win-win for me. If each cell really turned out to have the same response to different faces, then I would send Rodrigo an expensive bottle of wine. If on the other hand, the prediction did not pan out, he would send me solace wine.

In search of an answer back in our lab at Caltech, Le Chang first mapped the preferred axis for a given cell using responses to the 2,000 faces. Then he generated, while still recording from the same cell, a range of faces that could all be placed on an axis perpendicular to the cell's preferred axis. Remarkably, all these faces elicited exactly the same response in the cell. The next week Rodrigo received an exquisite bottle of Cabernet.

The finding proved that face cells are not encoding the identities of specific individuals in the IT cortex. Instead they are performing an axis projection, a much more abstract computation.

An analogy can be made to color. Colors can be coded by specific names, such as periwinkle, celandine and azure. Alternatively, one can code colors by particular combinations of three simple numbers that represent the amount of red, green and blue that make up that color. In the latter scheme, a color cell performing a projection onto the red axis would fire electrical impulses, or spikes, proportional

to the amount of red in any color. Such a cell would fire at the same intensity for a brown or yellow color containing the same amount of red mixed in with other colors. Face cells use the same scheme, but instead of just three axes, there are 50. And instead of each axis coding the amount of red, green or blue, each axis codes the amount of deviation of the shape or appearance of any given face from an average face.

It would seem then that the Jennifer Aniston cells do not exist, at least not in the IT cortex. But single neurons responding selectively to specific familiar individuals may still be at work in a part of the brain that processes the output of face cells. Memory storage regions—the hippocampus and surrounding areas—may contain cells that help a person recognize someone from past experience, akin to the famed grandmother cells.

Facial recognition in the IT cortex thus rests on a set of about 50 numbers in total that represent the measurement of a face along a set of axes. And the discovery of this extremely simple code for face identity has major implications for our understanding of visual object representation. It is possible that all of the IT cortex might be organized along the same principles governing the face-patch system, with clusters of neurons encoding different sets of axes to represent an object. We are now conducting experiments to test this idea.

NEURAL ROSETTA STONE

IF YOU EVER GO to the British Museum, you will see an amazing artifact, the Rosetta stone, on which the same decree of Memphis is engraved in three different languages: Egyptian hieroglyphics, Demotic and Ancient Greek. Be-

cause philologists knew Ancient Greek, they could use the Rosetta stone to help decipher Egyptian hieroglyphics and Demotic. Similarly, faces, face patches and the IT cortex form a neural Rosetta stone—one that is still being deciphered. By showing pictures of faces to monkeys, we discovered face patches and learned how cells within these patches detect and identify faces. In turn, understanding coding principles in the face-patch network may one day lead to insight into the organization of the entire IT cortex, revealing the secret to how object identity more generally is encoded. Perhaps the IT cortex contains additional networks specialized for processing other types of objects—a whirring factory with multiple assembly lines.

I also hope that knowing the code for facial identity can help fulfill my college dream of discovering how we imagine curves. Now that we understand face patches, we can begin to train animals to imagine faces and explore how neural activity is shaped by the purely internal act of imagination. Lots of new questions arise. Does imagination reactivate the code for the imagined face in the face patches? Does it bring back even earlier representations of contours and shading that provide inputs to the face-patch system? We now have the tools to probe these questions and better understand how the brain sees objects, imagined or real.

Because almost all the brain's core behaviors—consciousness, visual memory, decision-making, language—require object interactions, a deep understanding of object perception will help us gain insight into the entire brain, not just the visual cortex. We are only starting to solve the enigma of the face. ■

MORE TO EXPLORE

The Code for Facial Identity in the Primate Brain. Le Chang and Doris Y. Tsao in *Cell*, Vol. 169, No. 6, pages 1013–1028; June 1, 2017.
How Do We Recognize a Face? Rodrigo Quian Quiroga in *Cell*, Vol. 169, No. 6, pages 975–977; June 1, 2017.

FROM OUR ARCHIVES

The Face as Entryway to the Self. Christof Koch; Consciousness Redux, *Scientific American Mind*, January/February 2015.

scientificamerican.com/magazine/sa