



©ISTOCKPHOTO.COM/MATJAZ SLANIC

Reading Minds

Brain-decoding scientists move closer to discovering the keys to unlock the brain.

By Leslie Mertz

When you see or think about an object, your brain engages in a unique pattern of activity tied specifically to that object. That's how you know a cat is a cat, and not a dog or a house or a cloud. Using functional magnetic resonance imaging (fMRI) and other techniques, scientists are not only able to measure those activity patterns but are also deciphering what each pattern means. Essentially, they are beginning to read minds.

That doesn't mean we all have to start wearing aluminum-foil hats to keep our thoughts to ourselves, but it does mean that scientists are creating an increasingly better instruction manual for the brain and its amazing processing capabilities. Such a manual can then be used to begin developing a brain-computer interface device that may help someone with a motor dysfunction to operate an exoskeleton for mobility or devising a system that can bypass a blind person's nonfunctional eyes and transmit images directly to the brain for translation into sight. It may also aid in the evolution of devices that permit faster and better recovery from stroke and other brain injuries or technologies for earlier diagnosis of such neural disorders as autism.

Digital Object Identifier 10.1109/MPUL.2017.2729412
Date of publication: 26 September 2017



FIGURE 1 John-Dylan Haynes of the Bernstein Center for Computational Neuroscience at the Charité–Universitätsmedizin. His research group is using fMRI to “find out what kinds of thoughts you can read out and what the principal limitations are.” (Photo courtesy of Bernstein Center for Computational Neuroscience.)



FIGURE 2 Marcel van Gerven, associate professor and principal investigator at Radboud University’s Donders Institute for Brain, Cognition, and Behavior, is developing advanced computer models for decoding brain activity. (Photo courtesy of Roy Soetekouw.)

Decrypting Brain Patterns

The primary methodology for studying brain-activity patterns is fMRI. It works like this. When the neurons in a part of the brain become active, the blood carries oxygenated hemoglobin to those areas. Tiny shifts in magnetic properties accompany this uptick in hemoglobin oxygenation, and fMRI is able to pick up those shifts. “So fMRI is actually measuring brain

activity indirectly using changes in the magnetization of blood,” explains John-Dylan Haynes, Ph.D., a professor with the Bernstein Center for Computational Neuroscience at Berlin’s Charité–Universitätsmedizin who specializes in cognitive neuroimaging (Figure 1).

A typical fMRI scan captures a 2–3-mm² voxel of brain area at a time, and each voxel can contain up to a million neurons. “We cannot resolve to the level of a single neuron or even a single blood vessel, so we’re measuring more of an aggregate blood signal within these voxels,” Haynes says. The temporal resolution of fMRI is also inexact because hemoglobin magnetization is gradual, building up and dropping down over several seconds. This smears the measurement over time, so fMRI measurements cannot precisely note the time of brain-signal onset.

Given that the resolution in space and time of fMRI is not perfect, why do researchers use it? The answer is very simple, Haynes says. “If you want to get a good spatial-resolution image of human brain activity, the only way to do this noninvasively is with fMRI, and that is why people have used it and continue to use it so much.” (A widely publicized research paper released in July 2016 [1] noted problems with a statistical correction used in some fMRI studies, and, although many media outlets interpreted the paper to suggest the invalidation of tens of thousands of studies and years of research, the paper’s coauthors have since noted that the statistical bug “only had a minor impact on the false positive rate” [2].)

In about the mid-2000s, researchers, including Haynes, began exploring brain-activity patterns using fMRI. “Part of the work that has been done over the last ten to 12 years internationally, including by our group, was just to find out what kinds of thoughts you can read out and what the principal limitations are,” he notes. The focus of most of those studies centers on reading the brain-activity patterns while a subject is looking at an image, so that researchers can generate a dictionary of brain patterns and their meanings for each individual subject.

Decoding Brain Activity

One researcher developing advanced computer models for decoding brain activity is Marcel van Gerven, Ph.D., associate professor and principal investigator at Radboud University’s Donders Institute for Brain, Cognition, and Behavior in Nijmegen, The Netherlands (Figure 2). His group is especially interested in exploring neural networks as computational models of human brain function and using the power of these models to improve decoding algorithms.

“On the computational side, these models are difficult to develop, but we have early, unpublished results showing that we can basically condition these models on brain function,” van Gerven says. “What happens is, we measure brain activity, the models observe this brain activity, and the models are able to make reconstructions based on that brain activity.”

One of the biggest challenges with the models is that they are built on fMRI data, which have inherent temporal limitations. “If something happens in my brain now, it could cause a change in blood oxygenation six seconds later, so we have these very slow measurements in fMRI while we are trying to reconstruct

what people are perceiving or imagining,” van Gerven describes. “For static stimuli, it’s doable. But the next steps—and we have been working on this—are to move toward more naturalistic stimuli, such as audiovisual stimuli, that are changing on a moment-to-moment basis.”

To continue down that path, van Gerven is amassing as much fMRI information as he can. “One of the things my group is focusing on is collecting huge amounts of data in individual subjects. In fact, we now have one participant who will be in the scanner for a [combined] total of 40 hours, with the objective of getting enough data to be able to estimate those models,” he says. “And the more data we have, the better those models become.”

Van Gerven’s earlier models were able to reconstruct observed images—even discerning different letters, such as an L versus an I—from fMRI data [3]. The group’s most recent models, which utilize neural-network techniques, tackle more complex objects [4]. “We have people perceive faces in the scanner, and we are able to make reconstructions of those faces,” he says. “That, for us, is now state of the art” (Figure 3).

While he and his group have been mainly interested in gaining insights into neuronal processing and the relationship between perception and imagery, they have just received a grant to apply some of that knowledge to help blind people see. “This is kind of the inverse of brain decoding, so this isn’t reading information from the human brain, but implanting information into the human brain,” van Gerven says. For this multipartner project, his group is working on computer modeling for an advanced system that transmits information from a head-mounted camera to be visually processed by the brain via an implantable electrode array [5]. This system builds on earlier work and takes it to a substantially elevated level of sophistication. He hopes that in five years or so “we will have the first prototype, which will be able to partially restore vision in blind people.”

Deep Thoughts

Now that researchers have clearly demonstrated they can decode brain activity patterns for observed objects, many are exploring whether they can do the same when a person is merely

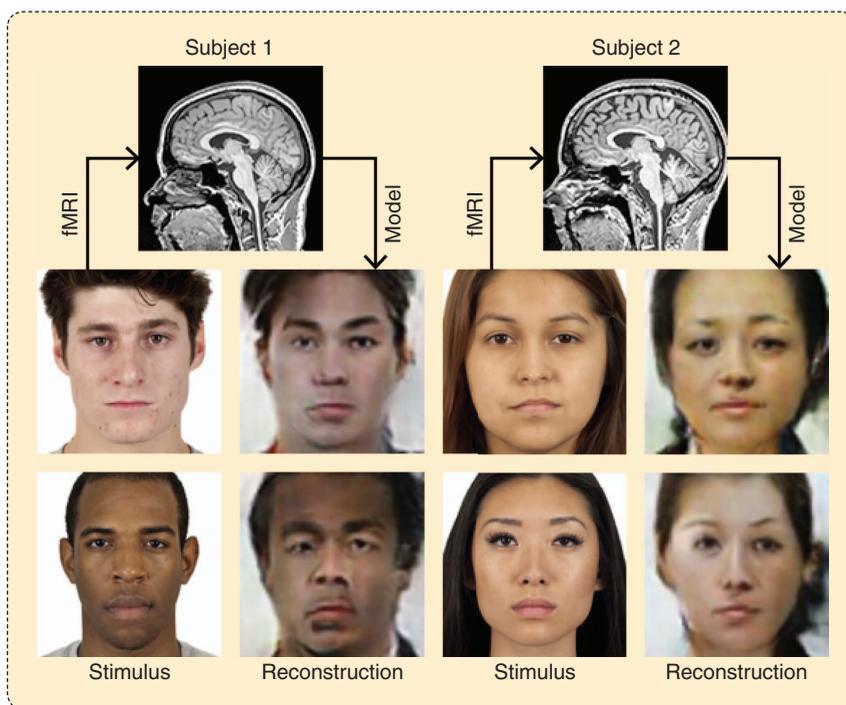


FIGURE 3 Van Gerven’s group is developing models that can build reconstructions of faces perceived by subjects. Here, two subjects view pictures of faces (stimulus), an fMRI scan captures the subjects’ brain activity, and the model translates that activity to generate reconstructed images (reconstruction). (Images courtesy of Marcel van Gerven and the Chicago Face Database, <http://faculty.chicagobooth.edu/bernd.wittenbrink/cfd/index.html>.)

imagining an object instead of actually looking at it. “Most people experience mental imagery as kind of a fuzzy, noisy, slippery approximation, but it still registers in our minds as a visual experience, and they happen all the time, pretty much nonstop,” says Thomas Naselaris, neuroscience researcher and assistant professor at the Medical University of South Carolina, Charleston (Figure 4).



FIGURE 4 Thomas Naselaris, neuroscience researcher and assistant professor at the Medical University of South Carolina. His group is modeling the stages of brain activity to decode what people are imagining. (Photo courtesy of Naselaris Lab.)

His interest in mental imagery is both basic and applied. Naselaris was intrigued by recent evidence showing that whether a person is observing something firsthand or imagining it, the same parts of the brain are active. “That seems a little odd,” he adds. “If your visual system evolved to help you interpret the sensory information that’s coming in so that you have a reliable report of things that are around you, it’s not totally clear why it would also be generating images of things that aren’t there. That’s a fascinating basic science question.”

To begin making sense of this riddle, Naselaris and his group took a closer look at the visual system, which he describes as a series of processing stages. “Each stage transforms visual information that’s coming from the eyes into a set of increasingly abstract

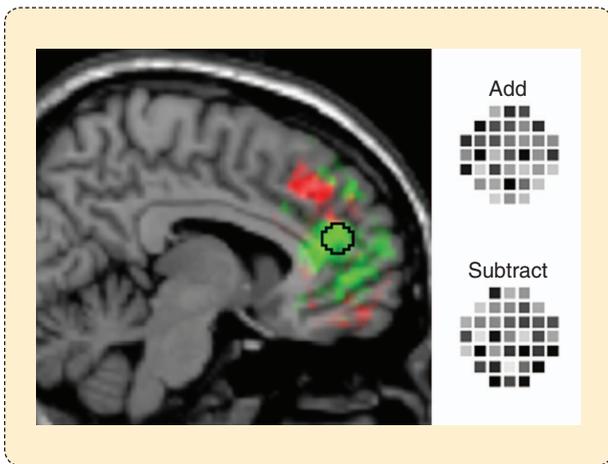


FIGURE 5 Haynes' research shows that it is possible to determine a subject's intentions—in this case, whether the person was preparing to perform an addition or a subtraction—by reading brain-activity patterns. Activity patterns in the green regions predicted covert intentions before the subject began to perform the calculation. The regions marked in red revealed intentions that were already being acted upon. (Photo courtesy of Bernstein Center for Computational Neuroscience.)

features that ultimately result in our ability to understand what we see or to extract the meaningful content of an image," he explains. "What we did was model what happens at each of the various stages and then use those tailored models to decode the picture people were imagining [6]. So, basically, we exploited the similarity between imagery and vision in order to access the imagery."

At this point, according to Naselaris, the models cannot together produce a pixel-by-pixel reconstruction of a mental image, but they can "summarize the textures, the edges, and some of the low-level features in the mental image." For instance, when a subject imagines one of a series of paintings by a certain artist, the model can use the fMRI data to surmise general features of the painting, but it cannot identify the specific painting.

Part of the reason for the model's inexact reconstruction is that, while brain-activity patterns are very similar between vision and mental imagery, bits of the patterns are accentuated differently. For example, the primary visual cortex (the part of the brain that receives information from the retina) is considerably more active when the subject is viewing rather than imagining an object, while activity in deeper brain regions is heightened when the subject is imagining the object rather than viewing it. At this point, Naselaris notes, the scientific community understands the primary visual cortex very well, but the deeper areas remain something of a mystery. He acknowledges, "That's one of the major outstanding challenges: learning visual

coding in brain areas that are most actively engaged during mental imagery."

To partially overcome that gap and refine the model's findings, Naselaris and his group ran the model's rather hazy results through an Internet image search to see whether those results were enough to identify the painting. And they found that the painting would indeed emerge near the top of the search. "This was a proof of principle," van Gersen says. "While the model is not doing complete reconstruction, it is definitely decoding a significant and interesting amount of information using brain activity."

He believes that the next step is one that affects the entire field of visual science: adopting innovative machine-learning (or "deep-learning") tools to generate a detailed map of brain activity. His group just published a paper [7] describing a way to map a large, complicated, and deep neural network and then regress the entire network onto single voxels in the brain, one voxel at a time—an approach that reveals intricate details about which layers or nodes in the deep neural network are most important.

"These machine-learning models, which are designed to solve engineering tasks, are turning out to be an excellent source of models for the brain and are actually quite like neural networks," Naselaris remarks, noting that he is participating in a conference group on cognitive computational neuroscience this fall. "The idea is to bring together neuroscientists, artificial intelligence [AI] researchers, and cognitive scientists to talk about how we can push AI forward using what we know about how the brain works, while leveraging what the AI scientists are doing to get a better understanding of the brain itself."

Looking Deeper Yet

As work on imaging continues, researchers are starting to ask whether it's possible to read what a person is feeling or thinking, or perhaps what the person is going to do next. The answer to all three is yes ... to a certain degree, according to Haynes. "You can read out different categories of thoughts from brain activity, but

you have to first learn the associations between the patterns of brain activity and the thoughts, and this is individual," he says. "Every person has [his or her] own way in which [to] code information. While it's not completely different from person to person, it is different enough that it's best if you learn how an individual brain itself codes the information, rather [than inferring] from someone else's brain."

To decode a person's thoughts is a huge task for many reasons, one of which is the sheer abundance of thoughts. Haynes provides a sample sentence from a Monty Python sketch: "This hovercraft is full of eels." He remarks, "It's a bizarre sentence. And if you were to build a universal mind-reading machine, which is a hypothetical device,



FIGURE 6 Kai Miller, a neurosurgery resident at Stanford University who is using ECoG in brain research. (Photo courtesy of Stanford University.)

you wouldn't necessarily have that sentence in your deciphering database." Haynes contends, "That just characterizes how difficult it is to come up with a system that can read out every thought."

Along this conjectural line of inquiry, Haynes is also interested in deciphering a person's intentions and resolving the role of brain activity in portending a person's actions (Figure 5). In other words, he asks, "When does the person feel [he or she] made a choice, and when did the brain give away the choice that person was going to make?" Through a number of studies, he and his research group found that brain-activity patterns foretell the outcomes several seconds *before* people think they make up their minds. "I may think I'm free to choose whether I move my left hand or right hand, or take an apple or orange. But even though I feel I haven't made up my mind, my brain may have already been biasing me in one direction or the other," he says.

In studying this and other provocative questions about how the brain works, researchers are also turning to approaches beyond fMRI. One of the most promising is electrocorticography (ECoG), in which electrodes are implanted directly on the surface of the brain. It is invasive and restricted mainly to consenting human subjects who have small strips of electrodes implanted for medical purposes, such as localizing the foci of epileptic seizures. Each electrode measures the electrical activity of the approximately half-million surrounding neurons and displays the average electrical activity of that population as it happens—without the delay seen in fMRI, claims Kai Miller, a neurosurgery resident at Stanford University who is using ECoG in his own brain research (Figure 6). Miller, who holds doctoral degrees in both physics and neurobiology as well as his medical degree, explains, "This very high-temporal precision lets you decode things very nicely, but it has its limitations in that you can only measure from those sites where small strips of electrodes are already implanted for clinical purposes."

Miller is combining signal-processing measures borrowed from electrical engineering and AI algorithms borrowed from computer science to gain insight into the types of computations performed by populations of neurons. "And as a byproduct of that work, we are able to start decoding the information content of different kinds of stimuli that we provide a patient, so we can look at the brain signal and start to predict what types of things people have seen and when they've seen it with very high precision," he says.

With what he learns, Miller ultimately hopes to develop implantable de-



FIGURE 7 Miller (shown here at right in surgery) hopes to develop implantable devices designed to promote brain plasticity and rehabilitation following brain injury, including stroke and tumor resection. (Photo courtesy of C.J. Kalkman, UMC Utrecht.)

VICES designed to promote brain plasticity and rehabilitation following a stroke, tumor resection, or other injury. "By plasticity, I'm talking about strengthening existing connections between brain areas. I want to see devices based on ECoG that can record activity in one brain region and use that to trigger paired stimulation of

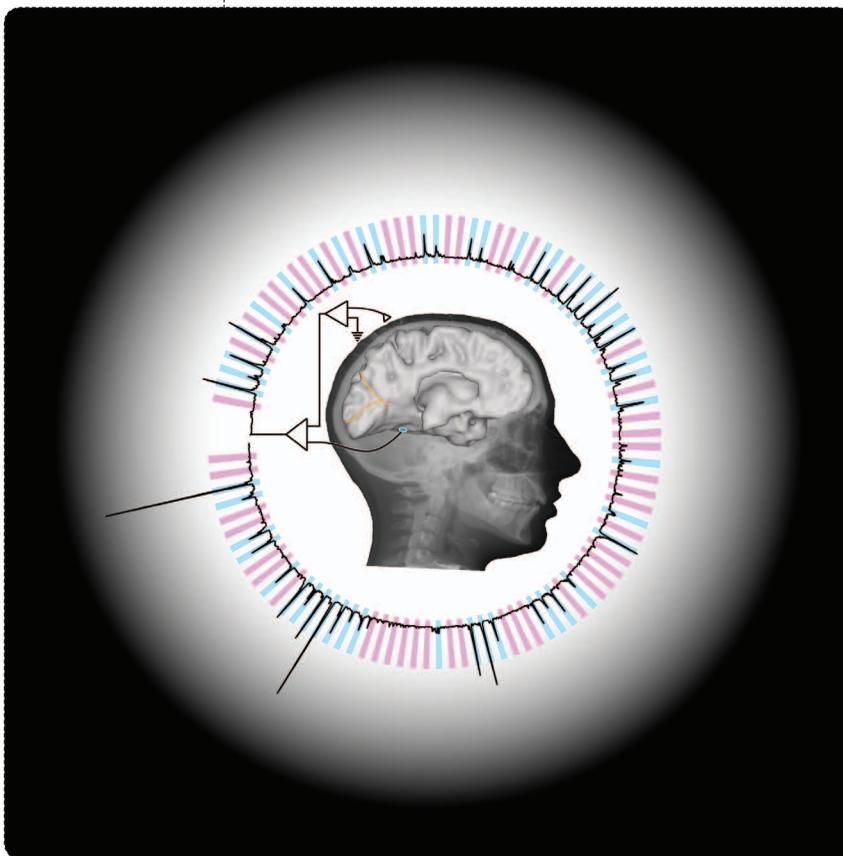


FIGURE 8 Using electrodes implanted in the temporal lobes of patients with epilepsy, researchers are using computational software to decode brain signals. This image shows the broadband response (black line) from an electrode (blue disk) as patients were shown images of faces (blue bars) and houses (pink bars) in 400-millisecond flashes. By combining these signals from around the brain, the researchers were able to accurately predict what patients saw with near-instantaneous precision. (Image courtesy of Kai Miller, Stanford University.)

multiple brain regions—both cortical [on the surface of the brain] and subcortical—to essentially trick the brain’s natural responses to induce plasticity and change,” he explains (Figure 7).

When asked how far along this project is, Miller responds, “It’s difficult to say because it depends on what it’s going to take to induce plasticity and we’re not really sure.” He and others had already shown that by pairing the brain activity involved in imagined movement or imagined speech to the movement of a cursor on a screen, patients can learn through operant conditioning to augment the activity in those brain areas within about ten minutes [8].

He continues to use ECoG to improve the ability to decode visual perception. “If I show people two broad classes of images, let’s say pictures of lots of faces and pictures of lots of houses, and I show those images a couple of seconds apart and in random order, I can spontaneously predict from the brain’s signals to within about 20 milliseconds and with 95% accuracy what the patient is seeing,” he says (Figure 8). He and his research group have also demonstrated that they can predict, with approximately the same accuracy, whether subjects are able to perceive or get a meaningful interpretation of what they’ve seen [9].

By taking the best of both worlds—fMRI’s more expansive spatial coverage of the brain and ECoG’s temporal detail but over a smaller area—Miller is obtaining an in-depth view of how and when different brain regions interact during visual perception. The timing is important because, while fMRI may show that perhaps five brain regions are active overall, ECoG discloses that the regions don’t become active all at once but in an extremely rapid-fire sequence. “By using the information about timing, I want to understand and perhaps generate new strategies that the brain might have for perceiving information content from the outside world after parts of that network have been destroyed due to injury.”

Putting Mind Reading to Work

The imagination runs wild when thinking about the potential applications of mind reading. Business people, for instance, are already talking about neural marketing and how they may one day be able to tap into consumers’ conscious and subconscious thoughts to trigger sales.

But what is actually possible? “If we look at what people have been claiming in the media, it gives the impression that we are already reading people’s minds, but I think that is kind of an overstatement at the moment,” van Gerven remarks. “We can do certain things, and we can’t do certain things. While we can read fMRI activity to reconstruct what somebody sees, for instance, this is a long way from being able to read someone’s mind, which includes all of a person’s beliefs, desires, and intentions.”

One very practical and often-discussed possibility for mind reading is lie detection. “Today, an fMRI lie detector works in the lab, but not perfectly,” says Haynes, noting that a shrewd subject can fool the lie detector. “At the same time, however, the other techniques we use habitually in the courtroom to decide

the truth are *really* flawed. Think about how a judge or someone on the jury uses intuition to ultimately believe one person and not the other.”

To make a reliable lie detector or any other mind-reading device, researchers need to collect much more subject data to fully understand the links between activity patterns and thoughts, Haynes continues. “Neural marketing, lie detectors, brain-computer interfaces, and all of these other applications are fascinating, but there are a lot of questions we have to ask about what we really want to know and how we can prove we’ve got the right information to develop them.”

Adds Haynes, “I’m very enthusiastic about the research field. In terms of getting this stuff into application, though, I think we are still not that far yet.”

Leslie Mertz (lmertz@nasw.org) is a freelance science, medical, and technical writer, author, and educator living in northern Michigan.

References

- [1] A. Eklund, T. E. Nichols, and H. Knutsson, “Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates,” *Proc. Nat. Acad. Sci.*, vol. 113, no. 28, pp. 7900–7905, July 12, 2016.
- [2] A. Eklund, T. Nichols, and H. Knutsson, “Reply to Cox et al. and Kessler et al.: Data and code sharing is the way forward for fMRI,” *Proc. Nat. Acad. Sci.*, vol. 114, no. 17, pp. E3374–E3375, Apr. 25, 2017.
- [3] S. Schoenmakers, M. Barth, T. Heskes, and M. van Gerven, “Linear reconstruction of perceived images from human brain activity,” *NeuroImage*, vol. 83, pp. 951–961, Dec. 2013.
- [4] Y. Güçlütürk, U. Güçlü, K. Seeliger, S. Bosch, R. van Lier, and M. van Gerven. (2017, May 19). Deep adversarial neural decoding. *arXiv*. [Online]. Available: <https://arxiv.org/abs/1705.07109>
- [5] Donders Institute for Brain, Cognition, and Behaviour. (2016, Nov. 18). STW funding to restore sight in the blind (press release). [Online]. Available: <http://www.ru.nl/donders/agenda-news/vm-news/restoring-sight-blind/>
- [6] T. Naselaris, C. A. Olman, D. E. Stansbury, K. Ugurbil, and J. A. Gallant, “A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes,” *NeuroImage*, vol. 105, pp. 215–228, Jan. 15, 2015.
- [7] G. St-Yves and T. Naselaris, “The feature-weighted receptive field: An interpretable encoding model for complex feature spaces,” *NeuroImage*, to be published.
- [8] K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, and R. P. N. Rao, “Cortical activity during motor execution, motor imagery, and imagery-based online feedback,” *Proc. Nat. Acad. Sci.*, vol. 107, no. 9, pp. 4430–4435, Mar. 2, 2010.
- [9] K. J. Miller, G. Schalk, D. Hermes, J. G. Ojemann, and R. P. N. Rao. (2016, Jan. 28). Spontaneous decoding of the timing and content of human object perception from cortical surface recordings reveals complementary information in the event-related potential and broadband spectral change. *PLoS Computational Biology*. [Online]. 12(1). Available: <https://doi.org/10.1371/journal.pcbi.1004660>

