The Atlantic

SCIENCE

Neuroscientists Have Discovered a Phenomenon That They Can't Explain

"Scientists are meant to know what's going on, but in this particular case, we are deeply confused."

ED YONG JUNE 9, 2021

ADAM MAIDA / THE ATLANTIC / SHUTTERSTOCK

Carl Schoonover and Andrew Fink are confused. As neuroscientists, they know that the brain must be flexible but not *too* flexible. It must rewire itself in the face of new experiences, but must also consistently represent the features of the external world. How? The relatively simple explanation found in neuroscience textbooks is that specific groups of neurons reliably fire when their owner smells a rose, sees a sunset, or hears a bell. These representations—these patterns of neural firing—presumably stay the same from one moment to the next. But as Schoonover, Fink, and others have found, they sometimes don't. They change—and to a confusing and unexpected extent.

Schoonover, Fink, and their colleagues from Columbia University allowed mice to sniff the same odors over several days and weeks, and recorded the activity of neurons in the rodents' piriform cortex—a brain region involved in identifying smells. At a given moment, each odor caused a distinctive group of neurons in this region to fire. But as time went on, the makeup of these groups slowly changed. Some neurons stopped responding to the smells; others started. After a month, each group was almost *completely different.* Put it this way: The neurons that represented the smell of an apple in May and those that represented the same smell in June were as different from each other as those that represent the smells of apples and grass at any one time.

This is, of course, just one study, of one brain region, in mice. But other scientists have shown that the same phenomenon, called representational drift, occurs in a variety of brain regions besides the piriform cortex. Its existence is clear; everything else is a mystery. Schoonover and Fink told me that they don't know why it happens, what it means, how the brain copes, or how much of the brain behaves in this way. How can animals possibly make any lasting sense of the world if their neural responses to that world are constantly in flux? If such flux is common, "there must be mechanisms in the brain that are undiscovered and even unimagined that allow it to keep up," Schoonover said. "Scientists are meant to know what's going on, but in this particular case, we are deeply confused. We expect it to take many years to iron out."

It had already taken years for Schoonover and Fink to even confirm that representational drift exists in the piriform cortex. They needed to develop surgical techniques for implanting electrodes into a mouse's brain and, crucially, keeping them in place for many weeks. Only then could they be sure that the drift they witnessed was really due to changes in the neurons, and not small movements of the electrodes themselves. They started working on this in 2014. By 2018, they were confident that they could get stable recordings. They then allowed implant-carrying mice to periodically inhale different odors.

The team showed that if a neuron in the piriform cortex reacts to a specific smell, the odds that it will still do so after a month are just one in 15. At any one time, the same number of neurons fires in response to each odor, but the identity of those neurons changes. Daily sniffs can slow the speed of that drift, but they don't eliminate it. Nor, bizarrely, does learning: If the mice associated a smell with a mild electric shock, the neurons representing that scent would still completely change even though the mice continued to avoid it. "The prevailing notion in the field has been that neuronal responses in sensory areas are stable over time," says Yaniv Ziv, a neurobiologist at the Weizmann Institute of Science who was not involved in the new study. "This shows that's not the case."

“There have been hints of this for at least 15 years,” across many parts of the brain, Schoonover told me. The hippocampus, for example, helps animals navigate their surroundings. It contains place cells—neurons that selectively fire when their owner enters specific locations. Walk from your bed to your front door, and different place cells will fire. But these preferences aren’t fixed: Ziv and others have now shown that the locations to which these cells are tuned can also drift over time.

In another experiment, Laura Driscoll, a neuroscientist who is now at Stanford, placed mice in a virtual T-shaped maze, and trained them to go either left or right. This simple task depends on the posterior parietal cortex, a brain region involved in spatial reasoning. Driscoll and her colleagues found that activity in this area also drifted: The neurons that fired when the mice ran the maze gradually changed, even though the rodents’ choices remained the same.

These results were surprising, but not overly so. The hippocampus is also involved in learning and short-term memory. You’d *expect* it to overwrite itself, and thus to continuously drift. “Up until now, observations of representational drift were confined to brain regions where we could tolerate it,” Schoonover said. The piriform cortex is different. It’s a sensory hub—a region that allows the brain to make sense of the stimuli around it. It ought to be stable: How else would smells ever be familiar? If representational drift can happen in the piriform cortex, it may be common throughout the brain.

It might be less common in other sensory hubs, such as the visual cortex, which processes information from the eyes. The neurons that respond to the *smell* of grass might change from month to month, but the ones that respond to the *sight* of grass seem to mostly stay the same. That might be because the visual cortex is highly organized. Adjacent groups of neurons tend to represent adjacent parts of the visual space in front of us, and this orderly mapping could constrain neural responses from drifting too far. But that might be true only for simple visual stimuli, such as lines or bars. Even in the visual cortex, Ziv found evidence of representational drift when mice watched the same movies over many days.

“We have a hunch that this should be the rule rather than the exception,” Schoonover said. “The onus now becomes finding the places where it *doesn’t* happen.” And in places where it does happen, “it’s the three F’s,” Fink added. “How fast does it go? How far does it get? And … how bad is it?”

How does the brain know what the nose is smelling or what the eyes are seeing, if the neural responses to smells and sights are continuously changing? One possibility is that it somehow corrects for drift. For example, parts of the brain that are connected to the piriform cortex might be able to gradually update their understanding of what the piriform’s neural activity means. The whole system changes, but it does so together.

Another possibility is that some high-level feature of the firing neurons stays the same, even as the specific active neurons change. As a simple analogy, “individuals in a population can change their mind while maintaining an overall consensus,” Timothy O’Leary, a neuroscientist at the University of Cambridge, told me. “The number of ways of representing the same signal in a large population is also large, so there's room for the neural code to move." Although some researchers have found signs of these stable, high-level patterns in other drifty parts of the brain, when Schoonover and Fink tried to do so in the piriform cortex, they couldn't. Neither they nor their colleagues can conclusively say how the brain copes with representational drift. They're also unsure *why* it happens at all.

Drift might simply be a nervous-system bug—a problem to be addressed. "The connections in many parts of the brain are being formed and broken down continually, and each neuron is itself continually recycling cellular material," O'Leary said. Perhaps a system like this—a gray, goopy version of the ship of Theseus—is destined to drift over time. But that idea "is a little weak," O'Leary told me. The nervous system *can* maintain precise and targeted connections, such as those between muscles and the nerves that control them. Drift doesn't seem inevitable.

Alternatively, drift might be beneficial. By constantly changing how existing information is stored, the nervous system might be better able to incorporate new material. "Information that's not continuously useful is forgotten, while information that continues to be useful is updated with the drift," says Driscoll, who is now testing this idea using artificial networks. "The more I've thought about drift, the more it makes sense that it's something we would see in the brain." Schoonover likes this idea too: "Our favored interpretation is that drift is a manifestation of learning," he told me. "It's not learning itself; it's the smoke that comes out of learning."

Schoonover and Fink compare the discovery of representational drift with the work of the astronomer Vera Rubin. In the 1970s, Rubin and her colleague Kent Ford noticed that some galaxies were spinning in unexpected ways that seemed to violate Newton's laws of motion. Her analysis of that data provided the first direct evidence for dark matter, which makes up most of the matter in the universe, but has never been observed. Similarly, drift indicates "that there's something else going on under the hood, and we don't know what that is yet," Schoonover said.

But the comparison between drift and Rubin's spinning galaxies fails in one important way. Rubin knew that she was onto something odd because she could compare her data against Newtonian mechanics—a solid and thoroughly described theory of physics. No such theory exists in neuroscience. The field has a very clear idea of how individual neurons work, but it gets much fuzzier when it comes to neuronal networks, entire brains, or the behavior of whole animals.

Consider the very idea that specific patterns of firing neurons can represent different smells, sights, or sounds. That connection seems simple enough—from the perspective of the experimenter, who exposed an animal to a stimulus and then looked for active neurons in its brain. But the brain itself has to work with just half of that equation, a bunch of active neurons, to make sense of what might have triggered that activity. "Just because we can decode that information doesn't mean the brain is doing that," says John Krakauer, a neuroscientist at Johns Hopkins University.

For that reason, Krakauer says that Schoonover and Fink's study, though "a technical tour de force," is also "very slightly straw-mannish." The idea of drift, he says, is surprising and exciting only when contrasted with the unsophisticated textbook idea of representations, which was never theoretically sound and was already being questioned. And that's a broader problem for the entire field, he told me. "Mainstream neuroscience relies on taking very specific methods and results and packaging them in a vague cloud of concepts that are only barely agreed upon by the field," he said. "In a lot of neuroscience, the premises remain unexamined, but everything else is impeccable."

Fink agrees that the idea of stable representations was never a theory—more "a tacit assumption," he said, and one that held "because it's simple." *How could it not be that way?* Well, it isn't. So now what?

"There's a real hunger in the field for new ideas," Fink told me, which is why, he thinks, he and Schoonover haven't yet faced the kind of vicious pushback that scientists with dogma-busting data tend to encounter. "People are really desperate for theories. The field is so immature conceptually that we're still at the point of collecting factlets, and we're not really in a position to rule anything out." Neuroscience's own representations of the brain still have plenty of room to drift.