

Neurons hide their memories in their imaginary fluctuations



Enlarge / This is your brain. Well, not *your* brain. Presumably your brain isn't being photographed at this moment.

reader comments

44 with 32 posters participating

Share this story

The brain is, at least to me, an enigma wrapped in a mystery. People who are smarter than I am—a list that encompasses most humans, dogs, and possibly some species of yeast—have worked out many aspects of the brain. But some seemingly basic things, like how we remember, are still understood only at a very vague level. Now, by investigating a mathematical model

of neural activity, researchers have found another possible mechanism to store and recall memories.

We know in detail how neurons function. Neurotransmitters, synapse firing, excitation, and suppression are all textbook knowledge. Indeed, we've abstracted these ideas to create blackbox algorithms to help us ruin people's lives by performing real-world tasks.

We also understand the brain at a higher, more structural, level: we know which bits of the brain are involved in processing different tasks. The vision system, for instance is mapped out in exquisite detail. Yet the intermediate level in between these two areas remains frustratingly vague. We know that a set of neurons might be involved in identifying vertical lines in our visual field, but we don't really understand how that recognition occurs.

Memory is hard

Likewise, we know that the brain can hold memories. We can even create and erase a memory in a mouse. But the details of how the memory is encoded are unclear. Our basic hypothesis is that a memory represents something that persists through time: a constant of sorts (we know that memories vary with recall, but they are still relatively constant). That means there should be something constant within the brain that holds the memory. But the brain is incredibly dynamic, and very little stays constant.

This is where the latest research comes in: abstract constants that may hold memories have been proposed.

So, what constants have the researchers found? Let's say that a group of six neurons is networked via interconnected synapses. The firing of any particular synapse is completely unpredictable. Likewise, its influence on its neighbors' activity is unpredictable. So, no single synapse or neuron

encodes the memory.

But hidden within all of that unpredictability is predictability that allows a neural network to be modeled with a relatively simple set of equations. These equations replicate the statistics of synapses firing very well (if they didn't, artificial neural networks probably wouldn't work).

A critical part of the equations is the weighting or influence of a synaptic input on a particular neuron. Each weighting varies with time randomly but can be strengthened or weakened due to learning and recall. To study this, the researchers examined the dynamical behavior of a network, focusing on the so-called fixed points (or set points).

Technically, you have to understand complex numbers to understand set points. But I have a short cut. The world of dynamics is divided into stable things (like planets orbiting the Sun), unstable things (like rocks balanced on pointy sticks), and things that are utterly unpredictable.

Memory is plastic

The neuron is a weird combination of stable and unpredictable. The neurons have firing rates and patterns that stay within certain bounds, but you can never know exactly when an individual neuron is going to fire. The researchers show that the characteristic that keeps the network stable does not store information for very long. However, the characteristic that drives unpredictability *does* store information, and it seems to be able to do so indefinitely.

The researchers demonstrated this by exposing their model to input stimulus, which they found changed the network's fluctuations. Furthermore, the longer the model was exposed to the stimulus, the stronger its influence was.

The individual pattern of firing was still unpredictable, and

there was no way to see the memory in the stimulus in any individual neuron or its firing behavior. Yet it was still there, hidden in the network's global behavior.

Further analysis shows that, in terms of the dynamics, there is a big difference between this way memory is encoded and previous models. In previous models, memory is a fixed point that corresponds to a particular pattern of neural firing. In this model, memory is a shape. It could be a 2D shape on a plane, as the researchers found in their model. But the dimensionality of the shape could be much larger, allowing very complicated memories to be encoded.

In a 2D model, the neuron-firing behavior follows a limit cycle, meaning that the pattern continuously changes through a range of states that eventually repeats itself, though this is only evident during recall.

Another interesting aspect of the model is that recall has an effect on the memory. Memories recalled by a similar stimulus get weaker in some cases, while in others they are strengthened.

Where to from here?

The researchers go on to suggest that evidence for their model might be found in biological systems. It should be possible to find invariant shapes in neuronal connectivity. However, I imagine that this is not an easy search to conduct. A simpler test is that there should be asymmetry in the strength in the connections between two neurons during learning. That asymmetry should change between learning and rest.

So, yes, in principle the model is testable. But it looks like those tests will be very difficult. We may be waiting a long time to get some results one way or another.

Nature Communications, 2019, DOI: [10.1038/s41467-019-12306-2](https://doi.org/10.1038/s41467-019-12306-2)

(About DOIIs)

<https://arstechnica.com/?p=1588143>