

Designing artificial brains can help us learn more about real ones

February 14 2020, by Blake Richards



Credit: Jessica Bryant from Pexels

Despite billions of dollars spent and decades of research, computation in

the human brain remains largely a mystery. Meanwhile, we have made great strides in the development of artificial neural networks, which are designed to loosely mimic how brains compute. We have learned a lot about the nature of neural computation from these artificial brains and it's time to take what we've learned and apply it back to the biological ones.

[Neurological diseases are on the rise worldwide](#), making a better understanding of computation in the brain a pressing problem. Given the ability of modern artificial neural networks to solve complex problems, a framework for neuroscience guided by machine learning insights may unlock valuable secrets about our own brains and how they can malfunction.

Our thoughts and behaviors are generated by computations that take place in our brains. To effectively treat neurological disorders that alter our thoughts and behaviors, like schizophrenia or depression, we likely have to understand how the computations in the brain go wrong.

However, understanding neural computation has proven to be an immensely difficult challenge. When neuroscientists record activity in the brain, it is [often indecipherable](#).

In a paper published in *Nature Neuroscience*, my co-authors and I argue that the lessons we have learned from artificial neural networks can [guide us down the right path of understanding the brain as a computational system rather than as a collection of indecipherable cells](#).

Brain network models

Artificial neural networks are computational models that loosely mimic the integration and activation properties of real neurons. They have become ubiquitous in the field of artificial intelligence.

To construct artificial neural networks, you start by first designing the network architecture: how the different components of the network are connected to one another. Then, you define the learning goal for the architecture, such as "learn to predict what you're going to see next." Then, you define a rule that tells the network how to change in order to achieve that goal using the data it receives.

What you do not do is specify how each neuron in the network is going to function. You leave it up to the network to determine how each neuron should function to best accomplish the task. I believe the development of the brain is probably the product of a similar process, [both on an evolutionary timescale and at the timescale of an individual learning within their lifetime.](#)

Assigning neuron roles

This calls into question the usefulness of trying to determine the functions of individual neurons in the brain, when it is possible that these neurons are the result of an optimization process much like what we see with artificial neural networks.

The different components of artificial neural networks are often very hard to understand. There's no simple verbal or simple mathematical description that explains exactly what they do.

In our paper, we propose that the same holds true for the brain, and so we have to move away from trying to understand the role of each neuron in the brain and instead look at the brain's architecture, that is its network structure; the optimization goals, either at the evolutionary timescale or within the person's lifetime; and the rules by which the brain updates itself—either over generations or within a lifetime—to meet those goals. By defining these three components, we may get a much better understanding of how the brain works than by trying to state

what each neuron does.

Optimizing frameworks

One successful application of this approach has shown that the dopamine releasing neurons in the brain appear to encode information about unexpected rewards, e.g. unexpected delivery of some food. This sort of signal, called a reward prediction error, is often used to train artificial neural networks to maximize the rewards they get.

For example, by programming an artificial neural network to interpret points received in a video game as a reward, you can use reward prediction errors to train the network how to play the video game. In the real brain, as in the artificial neural networks, even if we don't understand what each individual signal means, we can understand the role of these neurons and the neurons that receive their signals in relation to the learning goal of maximizing rewards.

While current theories in systems neuroscience are beautiful and insightful, I believe a cohesive framework founded in the way in which evolution and learning shape our brain could fill in a lot the blanks we have been struggling with.

To make progress in systems neuroscience, it will take both bottom-up descriptive work, such as tracing out the connections and gene expression patterns of cells in the brain, and top-down theoretical work, using artificial neural networks to understand learning goals and learning rules.

Given the ability of modern artificial neural networks to solve complex problems, a framework for systems neuroscience guided by machine learning insights may unlock valuable secrets about the human brain.

More information: Blake A. Richards et al. A deep learning framework for neuroscience, *Nature Neuroscience* (2019). [DOI: 10.1038/s41593-019-0520-2](https://doi.org/10.1038/s41593-019-0520-2)

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

This story is part of [Science X Dialog](#), where researchers can report findings from their published research articles. [Visit this page](#) for information about ScienceX Dialog and how to participate.

Provided by The Conversation

Citation: Designing artificial brains can help us learn more about real ones (2020, February 14) retrieved 22 August 2025 from <https://sciencex.com/news/2020-02-artificial-brains-real.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--