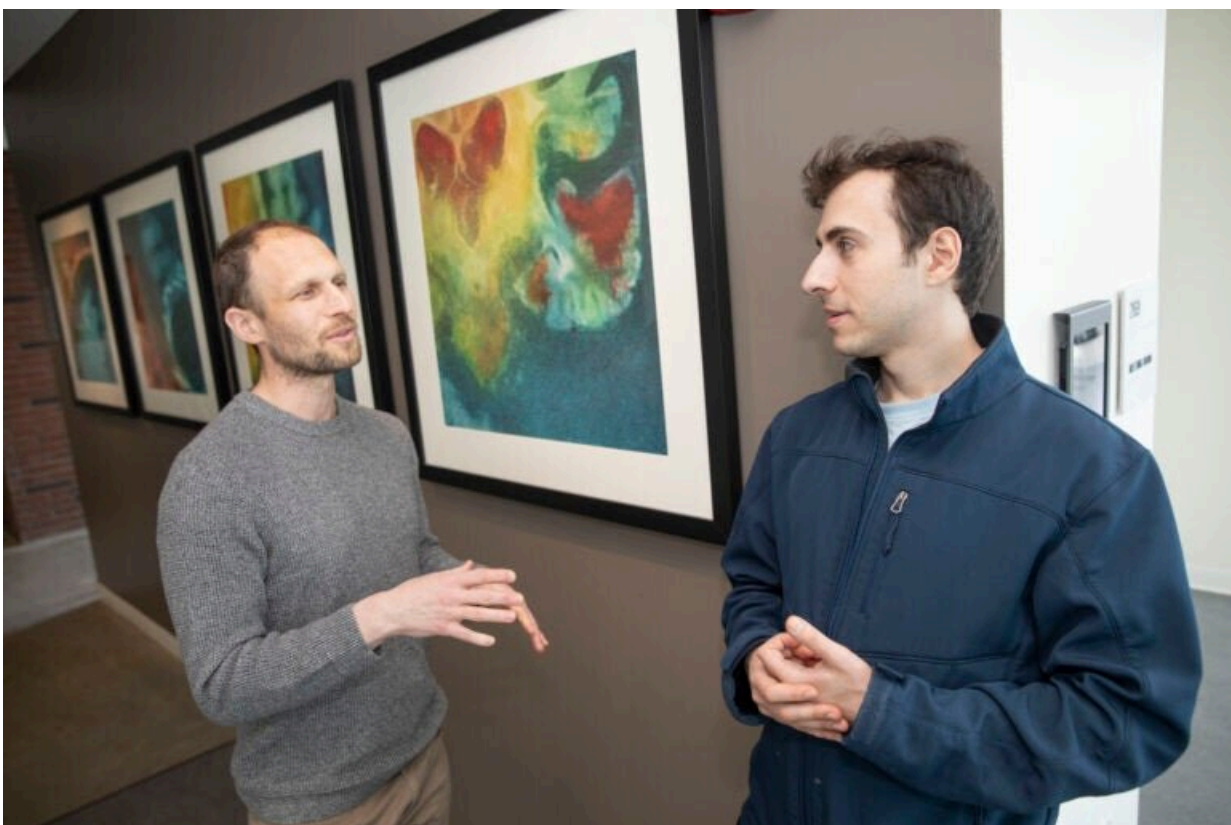


Making algorithms used in AI more human-like: Scientists use fMRI to test ideas about complex decision-making

July 13 2023, by Christy DeSmith



Professor Samuel Gershman and postdoc Momchil Tomov had volunteers play Atari-style video games while hooked up to fMRI scanners. Based on the results, they learned the directional flow of information was top-down during game play, the opposite of their original hypothesis. Credit: Kris Snibbe/Harvard Staff Photographer

How does the human brain navigate complex circumstances—say, driving through Harvard Square traffic at 5 p.m.?

One theory gaining support with psychologists and neuroscientists is that the [brain](#) creates causal models of the world that help with planning and execution. It's akin to running mental simulations to see which outcomes are good or bad. "You learn this internal model of the environment, which you can use to predict what will happen if you take different courses of action," explained Momchil Tomov, an associate in psychology Professor Samuel Gershman's Computational Cognitive Neuroscience Lab.

In recent decades, computer scientists have developed these ideas into a system dubbed Reinforcement Learning (or RL for short). Researchers such as Tomov who work at the intersection of psychology and technology have even introduced computational models that attempt to capture how RL plays out in the brain. In a new paper published in *Neuron*, Tomov and his co-authors used [functional magnetic resonance](#) (fMRI) to compare their algorithmic theory against real-world imaging.

Why craft algorithms that attempt to formalize human thinking and decision-making? "It's difficult to study [cognitive processes](#) without having a precise computational model that maps inputs to outputs," said Tomov, who earned his Ph.D. in neurobiology at Harvard in 2019 and worked with Gershman as a postdoc until 2021.

Researchers also hope their work leads to advances in RL, which can navigate complex environments and is considered one of the biggest success stories in artificial intelligence. It has, in fact, bested humans in realms including board and video games, but until recently has proven a somewhat slow learner. "Algorithms that are more human-like can perform better in certain domains than traditional machine-learning," Tomov said.

The group's experiment leans on the prior work of two of the study co-authors. Thomas Pouncy, another doctoral researcher in Gershman's lab, outlined in 2021 a more complex, theory-based RL system. A computational theory-based RL model was introduced in a subsequent paper by MIT postdoctoral researcher Pedro Tsividis. It proved much faster than previous iterations in learning new video games. In terms of speed, Tomov said, it's far closer to the human ability to pick up on such a task.

The whole process led the researchers to hypothesize on the neural architecture of human decision-making and learning. In the new study, the researchers tested their algorithm on 32 volunteers who played and eventually mastered Atari-style video games while hooked up to fMRI scanners, which measure the small changes in [blood flow](#) that come with brain activity.

As the researchers expected, this yielded evidence of activity theory-based models in the [prefrontal cortex](#) at the front of the brain with theory updates occurring in the posterior cortex, or back of the brain. Where their hypotheses—and their algorithm—diverged was in the details.

The researchers specifically expected to find evidence of theory-based models in the orbitofrontal cortex. Instead they found them in the [inferior frontal gyrus](#). This makes sense in hindsight, Tomov said, as previous research out of Gershman's lab found the inferior frontal gyrus involved with learning "causal rules that govern the world."

More surprises were found at the back of the brain, where the occipital cortex and the ventral pathway—both central to visual processing—appear to be involved when those models require updating. "Whenever you get surprising information that is inconsistent with your current theory, that's when we see not just an update signal in the ventral

pathway, but also, that's when the theory becomes activated in the inferior frontal gyrus," Tomov summarized.

Finally, fMRI scans revealed the directional flow of information in the brain. Tomov and his co-authors had hypothesized that information flows bottom-up. Instead, it seems to flow top-down during game play.

"It's almost as if it's coming from the model, stored somewhere in the prefrontal cortex, flowing down to the posterior visual regions," he said. "But then when there's a discrepancy—when an update happens—the pattern of information flow flips. Now information flows bottom-up, from posterior regions to frontal regions."

Tomov has been studying theory-based RL with Gershman for four years. Two years ago, he started applying these ideas to self-driving cars as a full-time employee with a Boston venture. "How do you get from here to the next intersection and make a left turn without hitting anyone?" he asked. "Basically, there's this internal [model](#) of the world with other drivers and predictions about what they're going to do."

More information: Momchil S. Tomov et al, The neural architecture of theory-based reinforcement learning, *Neuron* (2023). [DOI: 10.1016/j.neuron.2023.01.023](https://doi.org/10.1016/j.neuron.2023.01.023)

This story is published courtesy of the [Harvard Gazette](#), Harvard University's official newspaper. For additional university news, visit [Harvard.edu](https://www.harvard.edu).

Provided by Harvard University

Citation: Making algorithms used in AI more human-like: Scientists use fMRI to test ideas about complex decision-making (2023, July 13) retrieved 24 February 2024 from

<https://techxplore.com/news/2023-07-algorithms-ai-human-like-scientists-fmri.html>

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.