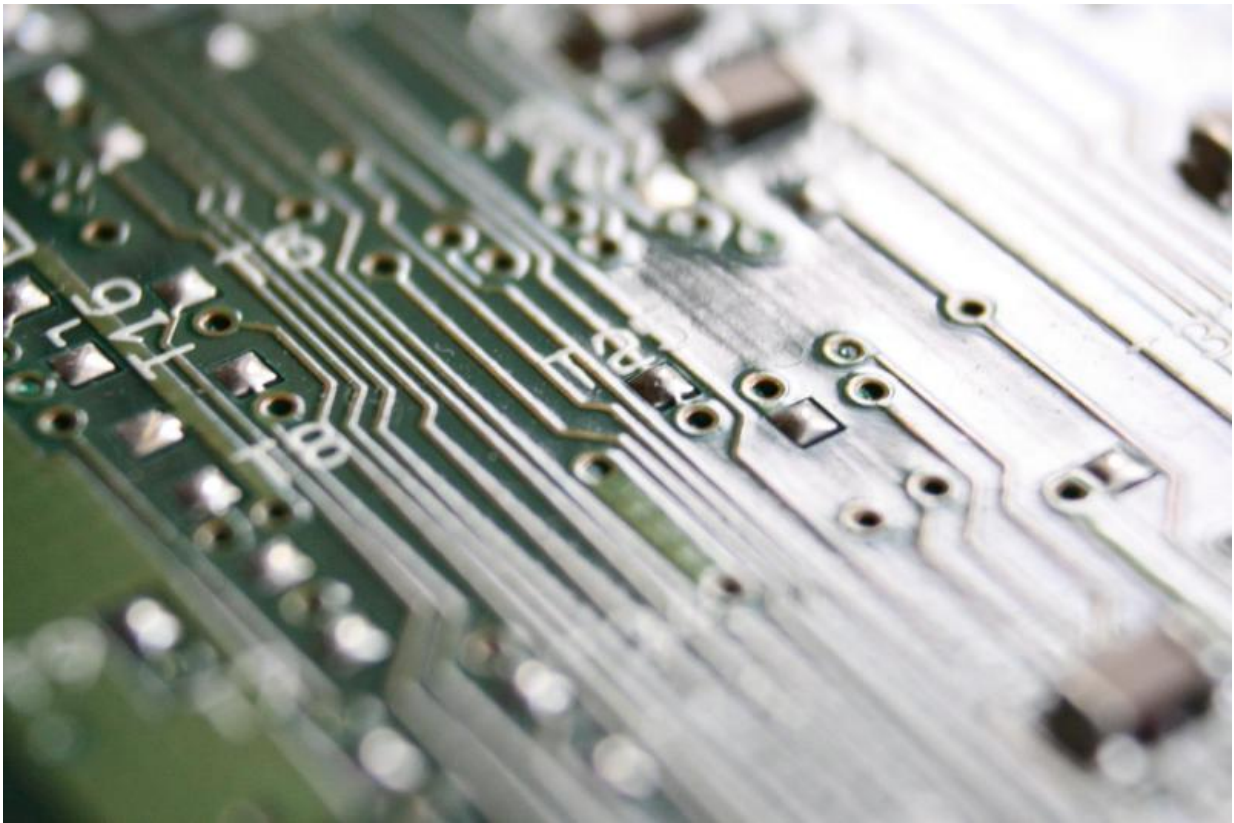# How insights into human learning can foster smarter artificial intelligence

June 14 2016

Credit: Public Domain

Recent breakthroughs in creating artificial systems that outplay humans in a diverse array of challenging games have their roots in neural networks inspired by information processing in the brain. In a Review

published June 14 in *Trends in Cognitive Sciences*, researchers from Google DeepMind and Stanford University update a theory originally developed to explain how humans and other animals learn - and highlight its potential importance as a framework to guide the development of agents with artificial intelligence.

First published in 1995 (*Psychol Rev.*, 102(3):419-57), the theory states that learning is the product of two complementary learning systems. The first system gradually acquires knowledge and skills from exposure to experiences, and the second stores specific experiences so that these can be replayed to allow their effective integration into the first system. The paper built on an earlier theory by influential British computational neuroscientist David Marr and on then-recent discoveries in neural network learning methods.

"The evidence seems compelling that the brain has these two kinds of learning systems, and the complementary learning systems theory explains how they complement each other to provide a powerful solution to a key learning problem that faces the brain," says Stanford Professor of Psychology James McClelland, lead author of the 1995 paper and senior author of the current Review.

The first system in the proposed theory, placed in the neocortex of the brain, was inspired by precursors of today's deep neural networks. As with today's deep networks, these systems contain several layers of neurons between input and output, and the knowledge in these networks is in their connections. Furthermore, their connections are gradually programmed by experience, giving rise to their ability to recognize objects, perceive speech, understand and produce language, and even to select optimal actions in game-playing and other settings where intelligent action depends on acquired knowledge.

Such systems face a dilemma when new information must be learned: If

large enough changes are made to the connections to force the new knowledge into the connections quickly, it will radically distort all of the other knowledge already stored in the connections.

"That's where the complementary learning system comes in," McClelland says. In humans and other mammals, this second system is located in a structure called the hippocampus. "By initially storing information about the new experience in the hippocampus, we make it available for immediate use and we also keep it around so that it can be replayed back to the cortex, interleaving it with ongoing experience and stored information from other relevant experiences." This two-system set-up therefore allows both immediate learning and also gradual integration into the structured knowledge representation in the neocortex.

"Components of the neural network architecture that succeeded in achieving human-level performance in a variety of computer games like Space Invaders and Breakout were inspired by complementary learning systems theory" says DeepMind cognitive neuroscientist Dharshan Kumaran, the first author of the Review. "As in the theory, these neural networks exploit a memory buffer akin to the hippocampus that stores recent episodes of game play and replays them in interleaved fashion. This greatly amplifies the use of actual game play experience and avoids the tendency for a particular local run of experience to dominate learning in the system."

Kumaran has collaborated both with McClelland and with DeepMind co-founder Demis Hassabis (also a co-author on the Review), in work that extended the role of the hippocampus as it was envisioned in the 1995 version of the complementary learning systems theory.

"In my view," says Hassabis, "the extended version of the complementary learning systems theory is likely to continue to provide a

framework for future research, not only in neuroscience but also in the quest to develop Artificial General Intelligence, our goal at Google DeepMind."

Provided by Cell Press