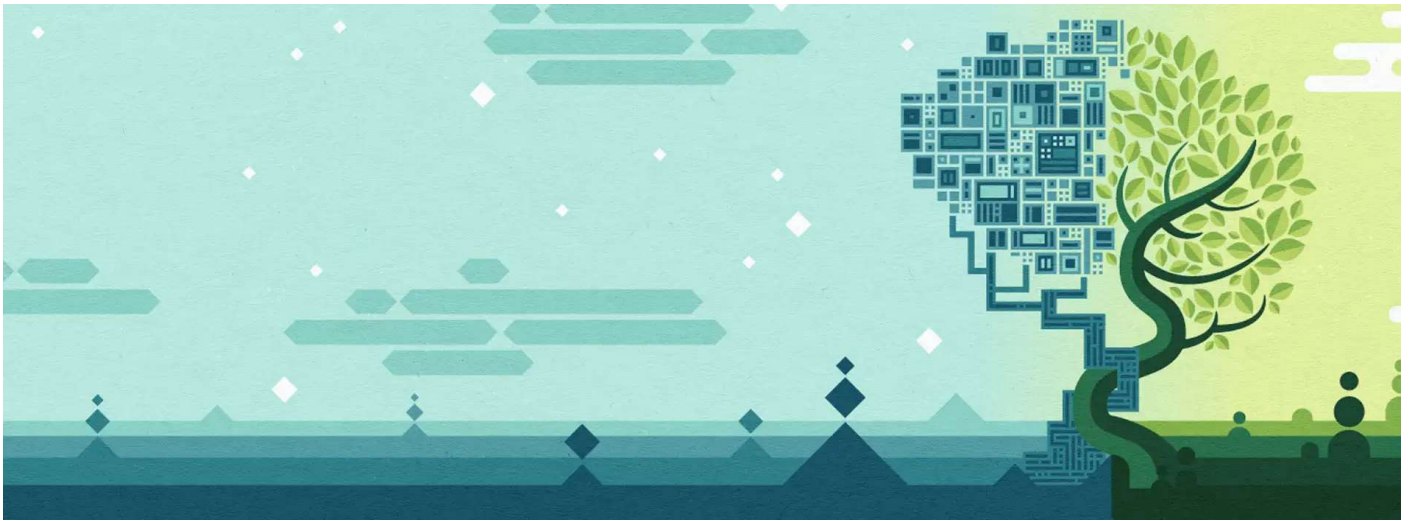


[< Blog](#)

BLOG POST  
RESEARCH

02 AUG 2017

## AI and Neuroscience: A virtuous circle

Recent progress in AI has been remarkable. Artificial systems now outperform expert humans at [Atari video games](#), the [ancient board game Go](#), and [high-stakes matches of heads-up poker](#). They can also produce [handwriting](#) and [speech](#) indistinguishable from those of humans, translate between multiple languages and even reformat your holiday snaps [in the style of Van Gogh](#) masterpieces.

These advances are attributed to several factors, including the application of new statistical approaches and the increased processing power of computers. But in [a recent Perspective in the journal Neuron](#), we argue that one often overlooked contribution is the use of ideas from experimental and theoretical neuroscience.

Psychology and neuroscience have played a key role in the history of AI. Founding figures such as [Donald Hebb](#), [Warren McCulloch](#), [Marvin Minsky](#) and [Geoff Hinton](#) were all originally motivated by a desire to understand how the brain works. In fact, throughout

“

**With so much at stake, the need for the field of neuroscience and AI to come together is now more urgent than ever before.**

At DeepMind, we argue that despite rapid progress in both fields, researchers should not lose sight of this vision. We urge researchers in neuroscience and AI to find a common language, allowing a free flow of knowledge that will allow continued progress on both fronts.

We believe that drawing inspiration from neuroscience in AI research is important for two reasons. First, neuroscience can help validate AI techniques that already exist. Put simply, if we discover one of our artificial algorithms mimics a function within the brain, it suggests our approach may be on the right track. Second, neuroscience can provide a rich source of inspiration for new types of algorithms and architectures to employ when building artificial brains. Traditional approaches to AI have historically been dominated by logic-based methods and theoretical mathematical models. We argue that neuroscience can complement these by identifying classes of biological computation that may be critical to cognitive function.

Take one recent example of a seminal finding in neuroscience: the discovery of offline experience “[replay](#)”. During sleep or quiet resting, biological brains “replay” temporal patterns of neuronal activity that were produced in an earlier active period. For example, when rats run through a maze, “place” cells activate as the animal moves around. During rest, the same sequence of neuronal activity is observed, as if the rats were mentally reimagining their past movements, and using them to optimise future behaviour. In fact,



'REPLAY' WAS A KEY ELEMENT OF DQN, A GENERAL-PURPOSE AGENT THAT IS ABLE TO CONTINUALLY ADAPT ITS BEHAVIOUR TO NEW ENVIRONMENTS

At first glance, it might seem counterintuitive to build an artificial agent that needs to 'sleep' – after all, they are supposed to grind away at a computational problem long after their programmers have gone to bed. But this principle was a key part of our [deep-Q network \(DQN\)](#), an algorithm that learnt to master a diverse range of Atari 2600 games to superhuman level with only the raw pixels and score as inputs. DQN mimics "experience replay", by storing a subset of training data that it reviews "offline", allowing it to learn anew from successes or failures that occurred in the past.

Successes like this give us confidence that neuroscience is already an important source of ideas for AI. But looking forward, we believe it will become indispensable in helping us tackle unsolved questions, such as those concerning efficient learning, understanding of the physical world, and imagination.

[Imagination](#) is a hugely important function for humans and animals, allowing us to plan for future scenarios without taking action; something that may come at a cost. Consider a simple example, such as planning a holiday. In order to do this we leverage our knowledge – or "model" – of the world and use it to project forward in time, evaluating



unveil the computational and systems mechanisms that underpin this kind of thinking, but much of this new understanding has yet to be incorporated into artificial models.



THE FIELDS OF NEUROSCIENCE AND ARTIFICIAL INTELLIGENCE HAVE A LONG AND INTERTWINED HISTORY

Another key challenge in contemporary AI research is known as transfer learning. To be able to deal effectively with novel situations, artificial agents need the ability to build on existing knowledge to make sensible decisions. Humans are already good at this – an individual who can drive a car, use a laptop or chair a meeting are usually able to cope even when confronted by an unfamiliar vehicle, operating system or social situation.

Researchers are now starting to take the first steps towards understanding how this might be possible in artificial systems. For example, a new class of network architecture known as a “[progressive network](#)” can use knowledge learned in one video game to learn another. The same architecture has also been shown to transfer knowledge from a simulated robotic arm to a real-world arm, massively reducing the training time. Intriguingly, these networks bear some similarities to [models of sequential task learning in humans](#). These tantalising links suggest that there are great opportunities for future AI research to learn from work in neuroscience.

approaches in contemporary AI research. Although the original idea came from theories of animal learning in psychology, it was developed and elaborated by machine learning researchers. These later ideas fed back into neuroscience to help us understand neurophysiological phenomena, such as the [firing properties of dopamine neurons](#) in the mammalian basal ganglia.

This back and forth is essential if both fields are to continue to build on each other's insights, creating a virtuous circle whereby AI researchers use ideas from neuroscience to build new technology, and neuroscientists learn from the behaviour of artificial agents to better interpret biological brains. Indeed, this cycle will likely accelerate thanks to recent advances, such as optogenetics, that allow us to precisely measure and manipulate brain activity, yielding vast quantities of data that can be analysed with tools from machine learning.

We therefore believe distilling intelligence into algorithms and comparing them to the human brain is now vital. Not only could it bolster our quest to develop AI, a tool that we hope will [create new knowledge and push forward scientific discovery](#), but may also allow us to better understand what's going on inside our own heads. That could shed light on some of the most enduring mysteries in neuroscience, such as the nature of creativity, dreams and, perhaps one day, even consciousness. With so much at stake, the need for the field of neuroscience and AI to come together is now more urgent than ever before.

---

Download the [full paper here](#).

### Further reading

