Rethinking the brain’s reward system—and what creates human intelligence

In a paper published in Nature … DeepMind, Alphabet’s AI subsidiary, has once again used lessons from reinforcement learning to propose a new theory about the reward mechanisms within our brains. The hypothesis, supported by initial experimental findings, could not only improve our understanding of mental health and motivation. It could also validate the current direction of AI research toward building more human-like general intelligence.

At a high level, reinforcement learning follows the insight derived from Pavlov’s dogs: it’s possible to teach an agent to master complex, novel tasks through only positive and negative feedback. An algorithm begins learning an assigned task by randomly predicting which action might earn it a reward. It then takes the action, observes the real reward, and adjusts its prediction based on the margin of error.

…

It turns out the brain’s reward system works in much the same way. … When a human or animal is about to perform an action, its dopamine neurons make a prediction about the expected reward.

…

What might it mean, for example, to have “pessimistic” and “optimistic” dopamine neurons? If the brain selectively listened to only one or the other, could it lead to chemical imbalances and induce depression?

Fundamentally, by further decoding processes in the brain, the results also shed light on what creates human intelligence.

Read the original post