Creating artificial consciousness

Ryota Kanai
Araya, Inc.

Abstract

The notion that consciousness is a natural phenomenon suggests that consciousness is subject to a set of yet to be uncovered laws of nature. However, conditions in which consciousness occurs remain elusive. Here, we review existing literature in psychology and neuroscience to discern possible functions of consciousness. Drawing upon empirical literature, we propose that a key function of consciousness might be the ability to internally generate sensory representations of events even when they are not happening at the present moment. Such counterfactual representations can be constructed by generative models that an agent learned through sensory-motor interactions with the environment. This ability endows an agent with a variety of cognitive functions related to consciousness such as intention, imagination, planning, short-term memory, attention, curiosity, and creativity, all of which contribute to behavioural flexibility and efficient learning. Using variational autoencoder (VAE) as an example, we illustrate how information generation and processing respectively corresponds to decoding and encoding, or data decompression and compression in artificial neural networks. In biological neural networks, information generation corresponds to top-down prediction, which is compatible with the common observation that feedback is more relevant for consciousness than feedforward processing. Taken together, the information generation theory provides new perspectives on the relationship between information and consciousness. Finally, I will discuss how artificial general intelligence (AGI) can be constructed based on the notion of information generation and possible links between general intelligence and consciousness.