The free-energy principle explains the brain
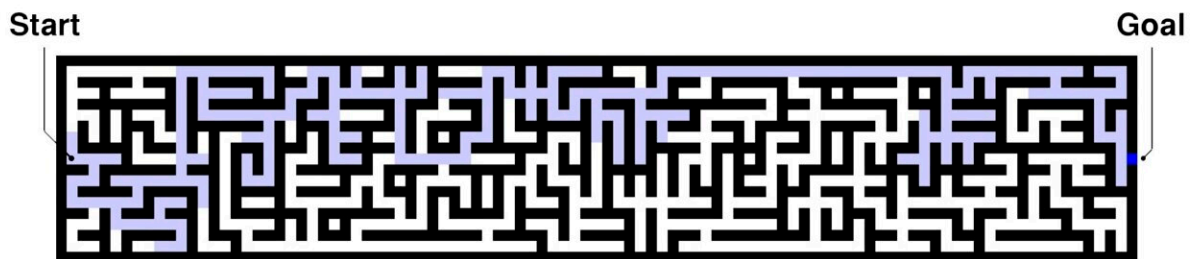
January 14 2022

General view of a solved maze. The maze comprises a discrete state space, wherein white and black cells indicate pathways and walls, respectively. The blue path is the trajectory. Starting from the left, the agent needs to reach the right edge of the maze within a certain amount of steps (time). The maze was solved following the free energy principle. Credit: RIKEN

The RIKEN Center for Brain Science (CBS) in Japan, along with colleagues, has shown that the free-energy principle can explain how neural networks are optimized for efficiency. Published in the scientific journal *Communications Biology*, the study first shows how the free-energy principle is the basis for any neural network that minimizes energy cost. Then, as proof of concept, it shows how an energy minimizing neural network can solve mazes. This finding will be useful for analyzing impaired brain function in thought disorders as well as for

generating optimized neural networks for artificial intelligences.

Biological optimization is a natural process that makes our bodies and behavior as efficient as possible. A behavioral example can be seen in the transition that cats make from running to galloping. Far from being random, the switch occurs precisely at the speed when the amount of energy it takes to gallop becomes less that it takes to run. In the brain, neural networks are optimized to allow efficient control of behavior and transmission of information, while still maintaining the ability to adapt and reconfigure to changing environments.

As with the simple cost/benefit calculation that can predict the speed that a cat will begin to gallop, researchers at RIKEN CBS are trying to discover the basic mathematical principles that underly how neural networks self-optimize. The free-energy principle follows a concept called Bayesian inference, which is the key. In this system, an agent is continually updated by new incoming sensory data, as well its own past outputs, or decisions. The researchers compared the free-energy principle with well-established rules that control how the strength of neural connections within a network can be altered by changes in sensory input.

"We were able to demonstrate that standard neural networks, which feature delayed modulation of Hebbian plasticity, perform planning and adaptive behavioral control by taking their previous 'decisions' into account," says first author and unit leader Takuya Isomura. "Importantly, they do so the same way that they would when following the free-energy principle."

Once they established that neural networks theoretically follow the free-energy principle, they tested the theory using simulations. The neural networks self-organized by changing the strength of their neural connections and associating past decisions with future outcomes. In this

case, the neural networks can be viewed as being governed by the free-energy principle, which allowed it to learn the correct route through a maze through trial and error in a statistically optimal manner.

These findings point toward a set of universal mathematical rules that describe how neural networks self-optimize. As Isomura explains, "Our findings guarantee that an arbitrary neural network can be cast as an agent that obeys the free-energy principle, providing a universal characterization for the brain." These rules, along with the researchers' new reverse engineering technique, can be used to study neural networks for decision-making in people with thought disorders such as schizophrenia and predict the aspects of their neural networks that have been altered.

Another practical use for these universal mathematical rules could be in the field of artificial intelligence, especially those that designers hope will be able to efficiently learn, predict, plan, and make decisions. "Our theory can dramatically reduce the complexity of designing self-learning neuromorphic hardware to perform various types of tasks, which will be important for a next-generation artificial intelligence," says Isomura.