How the Brain 'Constructs' the Outside World

Neural activity probes your physical surroundings to select just the information needed to survive and flourish

······························

By György Buzsáki on June 1, 2022

Credit: Stefania Infante

As a young course instructor in seminars for medical students, I faithfully taught neurophysiology by the book, enthusiastically explaining how the brain perceives the world and controls the body. Sensory stimuli from the eyes, ears, and such are converted to electrical signals and then transmitted to the relevant parts of the sensory cortex that process these inputs and induce perception. To initiate a movement, impulses from the motor cortex instruct the spinal cord neurons to produce muscular contraction.

Most students were happy with my textbook explanations of the brain's input-output mechanisms. Yet a minority—the clever ones—always asked a series of awkward questions. "Where in the brain does perception occur?" "What initiates a finger movement before cells in the motor cortex fire?" I would always dispatch their queries with a simple answer: "That all happens in the neocortex." Then I would skillfully change the subject or use a few obscure Latin terms that my students did not really understand but that seemed scientific enough so that my authoritative-sounding accounts temporarily satisfied them.

Like other young researchers, I began my investigation of the brain without worrying much whether this perception-action theoretical framework was right or wrong. I was happy for many years with my own progress and the spectacular discoveries that gradually evolved into what became known in the 1960s as the field of "neuroscience." Yet my inability to give satisfactory answers to the legitimate questions of my smartest students has haunted me ever since. I had to wrestle with the difficulty of trying to explain something that I didn't really understand.

Over the years I realized that this frustration was not uniquely my own. Many of my colleagues, whether they admitted it or not, felt the same way. There was a bright side, though, because these frustrations energized my career. They nudged me over the years to develop a perspective that provides an alternative description of how the brain interacts with the outside world.

The challenge for me and other neuroscientists involves the weighty question of what, exactly, is the mind. Ever since the time of Aristotle, thinkers have assumed that the soul or the mind is initially a blank slate, a tabula rasa on which experiences are painted. This view has influenced thinking in Christian and Persian philosophies, British empiricism and Marxist doctrine. In the past century it has also permeated psychology and cognitive science. This "outside-in" view portrays the mind as a tool for learning about the true nature of the world. The alternative view—one that has defined my research—asserts that the primary preoccupation of brain networks is to maintain their own internal dynamics and perpetually generate myriad nonsensical patterns of neural activity. When a seemingly random action offers a benefit to the organism's survival, the neuronal pattern leading to that action gains meaning. When an infant utters "te-te," the parent happily offers the baby "Teddy," so the sound "te-te" acquires the meaning of the Teddy bear. Recent progress in neuroscience has lent support to this framework.

DOES THE BRAIN "REPRESENT" THE WORLD?

Neuroscience inherited the blank slate framework millennia after early thinkers gave names like tabula rasa to mental operations. Even today we still search for neural mechanisms that might relate to their dreamed-up ideas. The dominance of the outside-in framework is illustrated by the outstanding discoveries of the legendary scientific duo David Hubel and Torsten Wiesel, who introduced single-neuronal recordings to study the visual system and were awarded the Nobel Prize in Physiology or Medicine in 1981. In their signature experiments, they recorded neural activity in animals while showing them images of various shapes. Moving lines, edges, light or dark areas, and other physical qualities elicited firing in different sets of neurons. The assumption was that neuronal computation starts with simple patterns that are synthesized into more complex ones. These features are then bound together somewhere in the brain to represent an object. No active participation is needed. The brain automatically performs this exercise.

The outside-in framework presumes that the brain's fundamental function is to perceive "signals" from the world and correctly interpret them. But if this assumption is true, an additional operation is needed to respond to these signals. Wedged between perceptual inputs and outputs resides a hypothetical central processor—which takes in sensory representations from the environment and makes decisions about what to do with them to perform the correct action.

So what exactly is the central processor in this outside-in paradigm? This poorly understood and speculative entity goes by various names—free will, homunculus, decision maker, executive function, intervening variables or simply just a "black box." It all depends on the experimenter's philosophical inclination and whether the mental operation in question is applied to the human brain, brains of other animals or computer models. Yet all these concepts refer to the same thing.

An implicit practical implication of the outside-in framework is that the next frontier for progress in contemporary neuroscience should be to find where the putative central processor resides in the brain and systematically elaborate the neuronal mechanisms of decision-making. Indeed, the physiology of decision-making has become one of the most popular focuses in contemporary neuroscience. Higher-order brain regions, such as the prefrontal cortex, have been postulated as the place where "all things come together" and "all outputs are initiated." When we look more closely, however, the outside-in framework does not hold together.

This approach cannot explain how photons falling on the retina are transformed into a recollection of a summer outing. The outside-in framework requires the artificial insertion of a human experimenter who observes this event [*see graphic below*]. The experimenter-in-

the-middle is needed because even if neurons change their firing patterns when receptors on sensory organs are stimulated—by light or sound, for instance—these changes do not intrinsically "represent" anything that can be absorbed and integrated by the brain. The neurons in the visual cortex that respond to the image of, say, a rose have no clue. They do not "see" the appearance of a flower. They simply generate electrical oscillations in response to inputs from other parts of the brain, including those arriving along multiple complex pathways from the retina.

In other words, neurons in sensory cortical areas and even in the hypothetical central processor cannot "see" events that happen in the world. There is no interpreter in the brain to assign meaning to these changes in neuronal firing patterns. Short of a magical homunculus watching the activities of all the neurons in the brain with the omniscience of the experimenter, the neurons that take this all in are unaware of the events that caused these changes in their firing patterns. Fluctuations in neuronal activity are meaningful only for the scientist who is in the privileged position of observing both events in the brain and events in the outside world and then comparing the two perspectives.

Outside In vs. Inside Out

....

The idea of the brain as a blank slate onto which experience is written has existed since ancient times—and persists today in modified form. Some neuroscientists have begun to question this theory because it requires a hard-to-justify assumption about the way we perceive and process events from the outside world—in particular, the need to involve a hypothetical "interpreter" to explain what is happening.

Outside-In Framework

A stimulus—the image of a flower—reaches the eyes, and the brain responds by causing neurons to fire. This theory is plausible only with the involvement of an "experimenter" to observe and establish a relation between the flower and the neuronal responses it induces. Absent the experimenter, neurons in the sensory cortex do not "see" the flower.
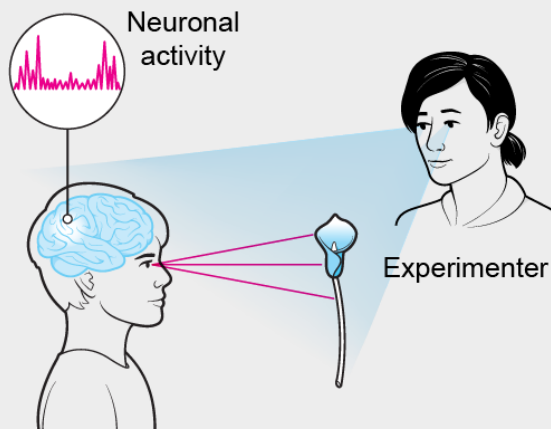
Inside-Out Framework

The alternative, inside-out theory does away with the experimenter. It presumes instead that we come to understand the external world by taking actions— moving a flower, for instance —to learn about an object. To accomplish this task, inputs from action-initiating neurons combine with sensory inputs to provide an understanding of the object's size, shape and other attributes. A meaningful picture arises, allowing the neurons to "see" the flower.

Neuronal activity

Experimenter

Neuronal activity

Action-initiating neurons

Sensory neurons
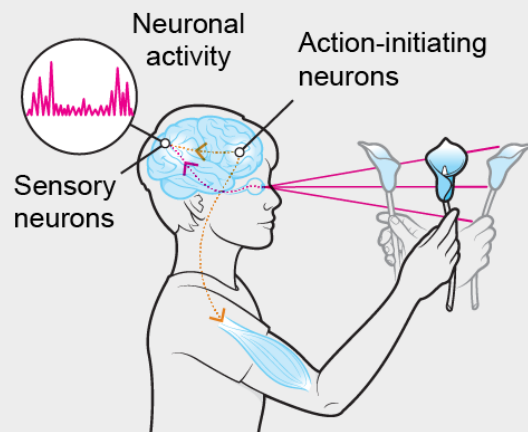
Credit: Brown Bird Design

PERCEPTION IS WHAT WE DO

....

Because neurons have no direct access to the outside world, they need a way to compare or "ground" their firing patterns to something else. The term "grounding" refers to the ability of the brain's circuits to assign meaning to changes in neuronal firing patterns that result from sensory inputs. They accomplish this task by relating this activity to something else. The "dah-dah-dit" Morse code pattern becomes meaningful only when it has previously been linked to the letter "G." In the brain, the only available source of a second opinion appears when we initiate some action.

We learn that sticks that look bent in water are not broken by moving them. Similarly, the distance between two trees and two mountain peaks may appear identical, but by moving around and shifting our perspective we learn the difference.

The outside-in framework follows a chain of events from perception to decision to action. In this model, neurons in dedicated sensory areas are "driven" by environmental signals and thus cannot relate their activity to something else. But the brain is not a serial processing unit; it does not proceed one by one through each of these steps. Instead any action a person takes involves the brain's motor areas informing the rest of the cerebral cortex about the action initiated—a message known as a corollary discharge.

Neuronal circuits that initiate an action dedicate themselves to two tasks. The first is to send a command to the muscles that control the eyes and other bodily sensors (the fingers and tongue, among others). These circuits orient bodily sensors in the optimal direction for in-depth investigation of the source of an input and enhance the brain's ability to identify the nature and location of initially ambiguous incoming signals from the senses.

The second task of these same action circuits involves sending notifications—the corollary discharges—to sensory and higher-order brain areas. Think of them as registered mail receipts. Neurons that initiate eye movement also notify visual sensory areas of the cortex about what is happening and disambiguate whether, say, a flower is moving in the wind or being handled by the person observing it.

This corollary message provides the second opinion sensory circuits need for grounding—a confirmation that "my own action is the agent of change." Similar corollary messages are sent to the rest of the brain when a person takes actions to investigate the flower and its relationship to oneself and other objects. Without such exploration, stimuli from the flower alone—the photons arriving on the retina connected to an inexperienced brain—would never become signals that furnish a meaningful description of the flower's size and shape. Perception then can be defined as what we *do*—not what we passively take in through our senses.

You can demonstrate a simple version of the corollary discharge mechanism. Cover one of your eyes with one hand and move the other eye gently from the side with the tip of your finger at about three times per second while reading this text. You will see immediately that the page is moving back and forth. By comparison, when you are reading or looking around the room, nothing seems to move. This constancy occurs because neurons that initiate eye movements to scan sentences also send a corollary signal to the visual system to indicate

whether the world or the eyeball is moving, thus stabilizing the perception of your surroundings.

LEARNING BY MATCHING

The contrast between outside-in and inside-out approaches becomes most striking when used to explain the mechanisms of learning. A tacit assumption of the blank slate model is that the complexity of the brain grows with the amount of experience. As we learn, the interactions of brain circuits should become increasingly more elaborate. In the inside-out framework, however, experience is not the main source of the brain's complexity.

Instead the brain organizes itself into a vast repertoire of preformed patterns of firing known as neuronal trajectories. This self-organized brain model can be likened to a dictionary filled initially with nonsensical words. New experience does not change the way these networks function—their overall activity level, for instance. Learning takes place, rather, through a process of matching the preexisting neuronal trajectories to events in the world.

To understand the matching process, we need to examine the advantages and constraints brain dynamics impose on experience. In its basic version, models of blank slate neuronal networks assume a collection of largely similar, randomly connected neurons. The presumption is that brain circuits are highly plastic and that any arbitrary input can alter the activity of neuronal circuits. We can see the fallacy of this approach by considering an example from the field of artificial intelligence. Classical AI research—particularly the branch known as connectionism, the basis for artificial neural networks—adheres to the outside-in, tabula rasa model. This prevailing view was perhaps most explicitly promoted in the 20th century by Alan Turing, the great pioneer of mind modeling: "Presumably the child brain is something like a notebook as one buys it from the stationer's," he wrote.

Artificial neural networks built to "write" inputs onto a neural circuit often fail because each new input inevitably modifies the circuit's connections and dynamics. The circuit is said to exhibit plasticity. But there is a pitfall. While constantly adjusting the connections in its networks when learning, the AI system, at an unpredictable point, can erase all stored memories—a bug known as catastrophic interference, an event a real brain never experiences.

The inside-out model, in contrast, suggests that self-organized brain networks should resist such perturbations. Yet they should also exhibit plasticity selectively when needed. The way the brain strikes this balance relates to vast differences in the connection strength of different groups of neurons. Connections among neurons exist on a continuum. Most neurons are only weakly connected to others, whereas a smaller subset retains robust links. The strongly

connected minority is always on the alert. It fires rapidly, shares information readily within its own group, and stubbornly resists any modifications to the neurons' circuitry. Because of the multitude of connections and their high communication speeds, these elite subnetworks, sometimes described as a "rich club," remain well informed about neuronal events throughout the brain.

The hard-working rich club makes up roughly 20 percent of the overall population of neurons, but it is in charge of nearly half of the brain's activity. In contrast to the rich club, most of the brain's neurons—the neural "poor club"—tend to fire slowly and are weakly connected to other neurons. But they are also highly plastic and able to physically alter the connection points between neurons, known as synapses.

Both rich and poor clubs are important for maintaining brain dynamics. Members of the ever ready rich club fire similarly in response to diverse experiences. They offer fast, good-enough solutions under most conditions. We can make good guesses about the unknown not because we remember it but because our brains always make a surmise about a new, unfamiliar event. Nothing is completely novel to the brain because it always relates the new to the old. It generalizes. Even an inexperienced brain has a vast reservoir of neuronal trajectories at the ready, offering opportunities to match events in the world to preexisting brain patterns without requiring substantial reconfiguring of connections. A brain that remakes itself constantly would be unable to adapt quickly to fast-changing events in the outside world.

But there also is a critical role for the plastic, slow-firing-rate neurons. These neurons come into play when something of importance to the organism is detected and needs to be recorded for future reference. They then go on to mobilize their vast reserve to capture subtle differences between one thing and another by changing the strength of some connections to other neurons. Children learn the meaning of the word "dog" after seeing various kinds of canines. When a youngster sees a sheep for the first time, they may say "dog." Only when the distinction matters—understanding the difference between a pet and livestock—will they learn to differentiate.

Imagining the Road Ahead

An experiment demonstrates that distinct sets of neurons fire—each set in a different order—when a rat is planning whether to take the left or right route to receive a reward.
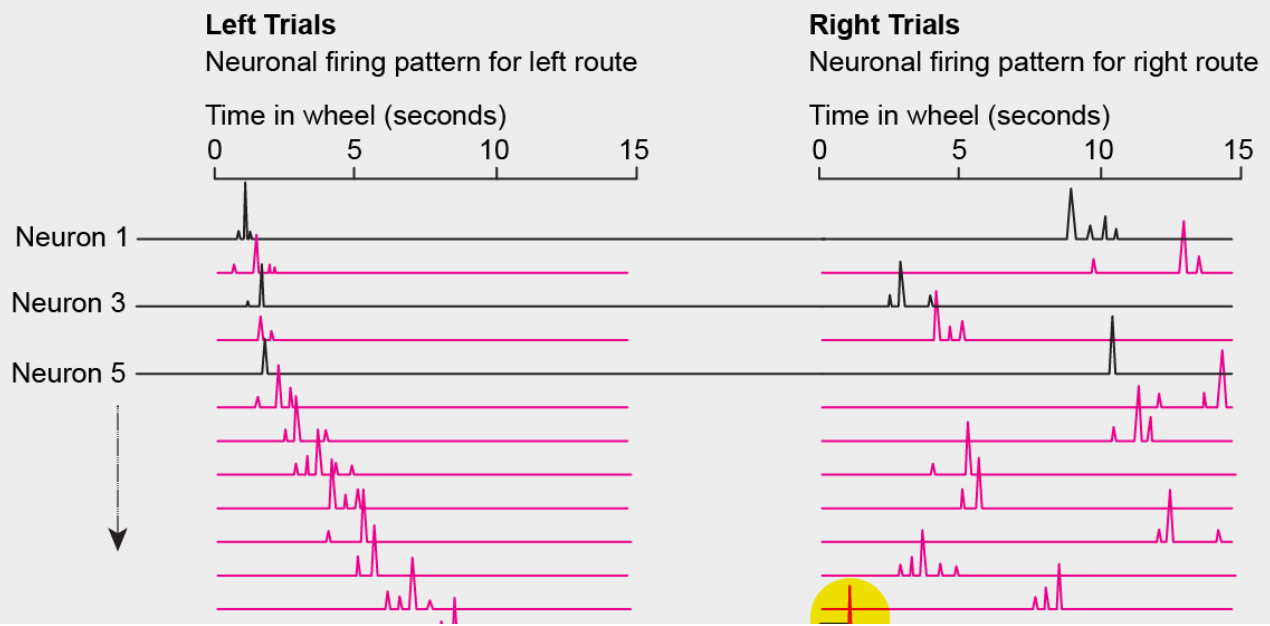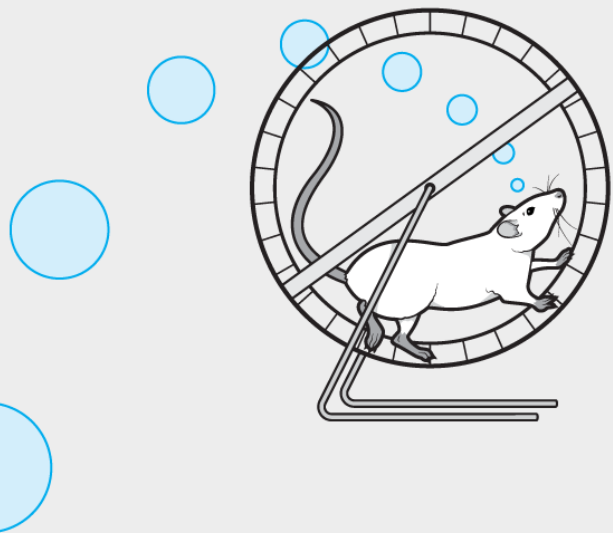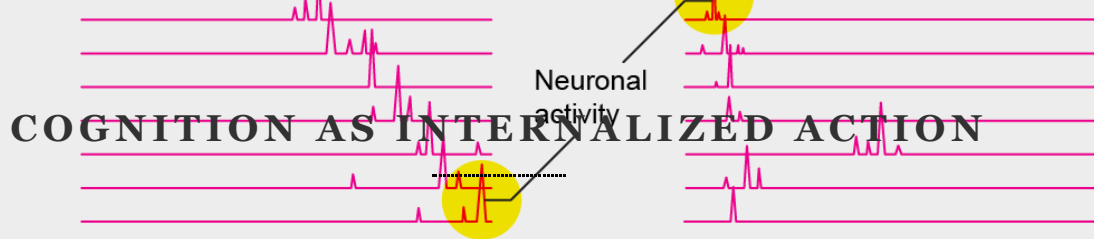
Experimental Setup

A running wheel is located at the entrance to a maze with two route options, both of which lead to a reward. The rat is free to choose a path through the maze after a 15-second run on the wheel. Neuronal firing patterns are recorded during both maze and wheel activity.

Left route
Right route
Reward
Running wheel
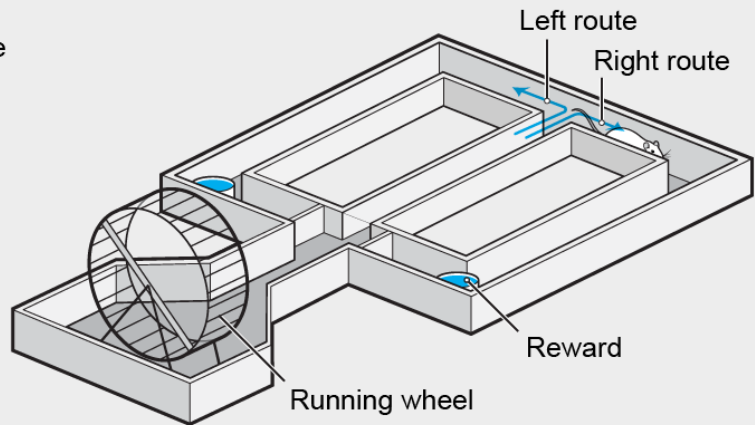
Results

Neuronal activity while the rat was running in the wheel predicted the direction it would take in the maze many seconds later, as if the animal were imagining the path to come. The "left trials" panel represents a sequence of neuronal firing that differs from the one for the right trials. When the "left" pattern occurred while the rat was in the wheel, it took the left route in the maze moments later.

Left Trials
Neuronal firing pattern for left route

Time in wheel (seconds)
0   5   10   15

Neuron 1
Neuron 3
Neuron 5

Right Trials
Neuronal firing pattern for right route

Time in wheel (seconds)
0   5   10   15

Neuronal activity

framework. Only decades after I started my work studying the self-organization of brain circuits and the rhythmic firing of neuronal populations in the hippocampus did I realize that the brain is more occupied with itself than with what is happening around it. This realization led to a whole new research agenda for my lab. Our experiments, along with findings from other groups, revealed that neurons devote most of their activity to sustaining the brain's perpetually varying internal states rather than being controlled by stimuli impinging on our senses.

During the course of natural selection, organisms adapt to the ecological niches in which they live and learn to predict the likely outcomes of their actions in those niches. As brain complexity increases, more intricate connections and neuronal computations insert themselves between motor outputs and sensory inputs. This investment enables the prediction of planned actions in more complex and changing environments and at lengthy time scales far in the future. More sophisticated brains also organize themselves to allow computations to continue when sensory inputs vanish temporarily and an animal's actions come to a halt. When you close your eyes, you still know where you are because a great deal of what defines "seeing" is rooted in brain activity itself. This disengaged mode of neuronal activity provides access to an internalized virtual world of vicarious or imagined experience and serves as a gateway to a variety of cognitive processes.

Let me offer an example of such a disengaged mode of brain operation from our work on the brain's temporal lobe, an area that includes the hippocampus, the nearby entorhinal cortex and related structures involved with multiple aspects of navigation (the tracking of direction, speed, distance traveled, environmental boundaries, and so on).

Our research builds on leading theories of the functions of the hippocampal system, such as the spectacular Nobel-winning discovery of John O'Keefe of University College London. O'Keefe found that firing of hippocampal neurons during navigation coincides with the spatial location of an animal. For that reason, these neurons are known as place cells.

When a rat walks through a maze, distinct assemblies of place cells become active in a sequential chain corresponding to where it is on its journey. From that observation, one can tentatively conclude that continually changing sensory inputs from the environment exercise control over the firing of neurons, in line with the outside-in model.

Yet other experiments, including in humans, show that these same networks are used for our internal worlds that keep track of personal memories, engage in planning and imagine future actions. If cognition is approached from an inside-out perspective, it becomes clear that navigation through either a physical space or a landscape that exists only in the imagination is processed by identical neural mechanisms.

Fifteen years ago my lab set about to explore the mechanisms of spatial navigation and memory in the hippocampus to contrast the outside-in and inside-out frameworks. In 2008 Eva Pastalkova, a postdoctoral fellow, and I trained rats to alternate between the left and right arms of a maze to find water. At the beginning of each traversal of the maze, the rat was required to run in a wheel for 15 seconds, which helped to ensure that memory alone of the maze routes, and not environmental and body-derived cues, allowed it to choose a particular arm of the maze. We reasoned that if hippocampal neurons "represent" places in the maze corridors and the wheel, as predicted by O'Keefe's spatial navigation theory, a few neurons should fire continuously at each spot whether the rat is in the corridors or the wheel. In contrast, if the neurons' firing is generated by internal brain mechanisms that can support both navigation and memory, the duration of neuronal firing should be similar at all locations, including inside the wheel.

The findings of these experiments defied outside-in explanations. Not a single neuron among the hundreds recorded fired continuously throughout the wheel running. Instead many neurons fired transiently one after the other in a continuous sequence.

Obviously these neurons could not be called place cells, because the animal's body was not displaced while at the single location of the running wheel. Moreover, the firing patterns of individual neurons in this neuronal trajectory could not be distinguished from neurons active when the rat was traversing the arms of the maze.

When we sorted individual trials according to the rat's future choice of left or right arms, the neuronal trajectories were uniquely different. The distinct trajectories eliminated the possibility that these neuronal sequences arose from counting steps, estimating muscular effort or some other undetected feedback stimuli from the body. Also, the unique neuronal trajectories allowed us to predict the animal's maze arm choice from the moment it entered the wheel and throughout wheel running, a period in which the rat had to keep in mind the previously visited arm. The animals needed to correctly choose the alternate maze arm each time to get their rewards [*see graphic above*].

These experiments lead us to the idea that the neuronal algorithms that we can use to walk to the supermarket govern internalized mental travel. Disengaged navigation takes us through

the series of events that make up personal recollections, known as episodic memories.

In truth, episodic memories are more than recollections of past events. They also let us look ahead to plan for the future. They function as a kind of "search engine" that allows us to probe both past and future. This realization also presages a broadening in nomenclature. These experiments show that progressions of place cell activity are internally generated as preconfigured sequences selected for each maze corridor. Same mechanism, multiple designations—so they can be termed place cells, memory cells or planning cells, depending on the circumstance.

Further support for the importance of disengaged circuit operations comes from "offline" brain activity when an animal is milling around doing nothing, consuming a reward or just sleeping. As a rat rests in the home cage after a maze exploration, its hippocampus generates brief, self-organized neuronal trajectories. These sharp wave ripples, as they are known, occur in 100-millisecond time windows and reactivate the same neurons that were firing during several seconds of maze running, recapitulating the neuronal sequences that occurred during maze traversals. Sharp wave-ripple sequences help to form our long-term memories and are essential to normal brain functioning. In fact, alteration of sharp wave-ripple events by experimental manipulations or disease results in serious memory impairment [*see graphic below*].
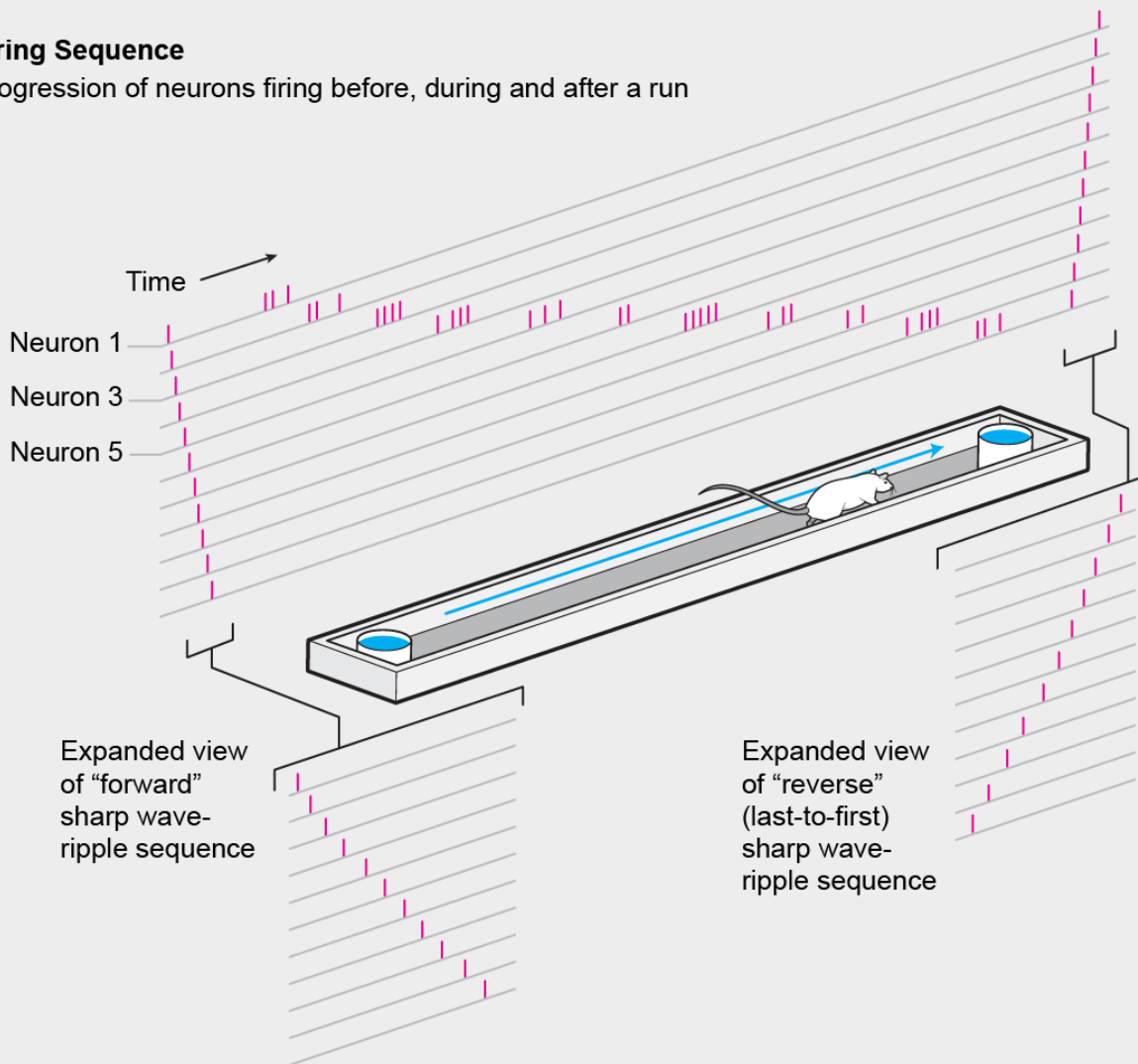
Clever experiments performed in human subjects and in animals over the past decade show that the time-compressed ripple events constitute an internalized trial-and-error process that subconsciously creates real or fictive alternatives for making decisions about an optimal strategy, constructing novel inferences and planning ahead for future actions without having to immediately test them by undertaking a real exploit. In this sense, our thoughts and plans are deferred actions, and disengaged brain activity is an active, essential brain operation. In contrast, the outside-in theory does not make any attempt to assign a role to the disengaged brain when it is at rest or even in the midst of sleep.

Rehearsal and Playback

·······························

A group of neurons fire before, during and after a rat runs a lap on an elevated track. Neurons firing rapidly at the beginning and end of the run (*insets*) are the same ones active during the run, constituting either a rehearsal or a playback (the latter in reverse) of the rat's trajectory. These early and late events are known as sharp wave ripples and enable a mental process that selects and remembers an optimal path.

Firing Sequence
Progression of neurons firing before, during and after a run

Time

Neuron 1

Neuron 3

Neuron 5

Expanded view of "forward" sharp wave-ripple sequence

Expanded view of "reverse" (last-to-first) sharp wave-ripple sequence

Credit: Brown Bird Design

THE MEANING OF INSIDE OUT

·······························

In addition to its theoretical implications, the inside-out approach has a number of practical applications. It may help in the search to find better diagnostic tools for brain disease. Current terminology often fails to describe accurately underlying biological mechanisms of mental and neurological illnesses. Psychiatrists are aware of the problem but have been

hindered by limited understanding of pathological mechanisms and their relation to symptoms and drug responses.

The inside-out theory should also be considered as an alternative to some of the most prevalent connectionist models for conducting AI research. A substitute for them might build models that maintain their own self-organized activity and that learn by "matching" rather than by continual adjustments to their circuitry. Machines constructed this way could disengage their operations from the inputs of electronic sensors and create novel forms of computation that resemble internal cognitive processes.

In real brains, neural processes that operate through disengagement from the senses go hand in hand with mechanisms that promote interactions with the surrounding world. All brains, simple or complex, use the same basic principles. Disengaged neural activity, calibrated simultaneously by outside experience, is the essence of cognition. I wish I had had this knowledge when my smart medical students asked their legitimate questions that I brushed off too quickly.

ABOUT THE AUTHOR(S)

György Buzsáki is a systems neuroscientist whose work has focused on the ways memories form and how brain rhythms segment neural information to support cognition. He was a co-recipient of the 2011 Brain Prize from the Lundbeck Foundation. Buzsáki is author most recently of *The Brain from Inside Out* (Oxford University Press, 2019).