NEUROSCIENCE

Neuroscience Readies for a Showdown Over Consciousness Ideas

By PHILIP BALL

March 6, 2019

To make headway on the mystery of consciousness, some researchers are trying a rigorous new way to test competing theories.

71

Neuroscientists are preparing to test their ideas about the origins of consciousness — the cognitive state of experiencing your own existence.

Ryan Garcia for Quanta Magazine

Some problems in science are so hard, we don't really know what meaningful questions to ask about them — or whether they are even truly solvable by science. Consciousness is one

of those: Some researchers think it is an illusion; others say it pervades everything. Some hope to see it reduced to the underlying biology of neurons firing; others say that it is an irreducibly holistic phenomenon.

The question of what kinds of physical systems are conscious "is one of the deepest, most fascinating problems in all of science," wrote the computer scientist Scott Aaronson of the University of Texas at Austin. "I don't know of any philosophical reason why [it] should be inherently unsolvable" — but "humans seem nowhere close to solving it."

Now a new project currently under review hopes to close in on some answers. It proposes to draw up a suite of experiments that will expose theories of consciousness to a merciless spotlight, in the hope of ruling out at least

some of them.

If all is approved and goes according to plan, the experiments could start this autumn. The initial aim is for the advocates of two leading theories to agree on a protocol that would put predictions of their ideas to the test. Similar scrutiny of other theories will then follow.

Whether or not this project, funded by the Templeton World Charity Foundation, narrows the options for how consciousness arises, it hopes to establish a new way to do science for difficult, contentious problems. Instead of each camp championing its own view and demolishing others, researchers will collaborate and agree to publish in advance how discriminating experiments might be conducted — and then respect the outcomes.

Dawid Potgieter, a senior program officer at the Templeton World Charity Foundation who is coordinating the endeavor, says that this is just the beginning of a sustained effort to winnow down theories of consciousness. He plans to set up several more of these "structured adversarial collaborations" over the next five years.

He is realistic about the prospects. "I don't think we are going to come to a single theory that tells us everything about consciousness," he said. "But if it were to take a hundred years to solve the mystery of consciousness, I hope we can cut it down to fifty."

A Workspace for Awareness

Philosophers have debated the nature of consciousness and whether it can inhere in

things other than humans for thousands of years, but in the modern era, pressing practical and moral implications make the need for answers more urgent. As artificial intelligence (AI) grows increasingly sophisticated, it might become impossible to tell whether one is dealing with a machine or a human merely by interacting with it — the classic Turing test. But would that mean AI deserves moral consideration?

Understanding consciousness also impinges on animal rights and welfare, and on a wide range of medical and legal questions about mental impairments. A group of more than 50 leading neuroscientists, psychologists, cognitive scientists and others recently called for greater recognition of the importance of research on this difficult subject. "Theories of consciousness need to be tested rigorously and

revised repeatedly amid the long process of accumulation of empirical evidence," the authors said, adding that "myths and speculative conjectures also need to be identified as such."

The cognitive scientist Stanislas Dehaene of the Collège de France in Paris is one of those behind the global workspace theory of

consciousness, which asserts that conscious behavior arises when sensory information collected in a cognitive "workspace" is broadcast to other brain centers.

Per Henning/NTNU

You can hardly do experiments on consciousness without having first defined it. But that's already difficult because we use the word in several ways. Humans are conscious beings, but we can lose consciousness, for example under anesthesia. We can say we are conscious of something — a strange noise coming out of our laptop, say. But in general, the quality of consciousness refers to a capacity to experience one's existence rather than just recording it or responding to stimuli like an automaton. Philosophers of mind often

refer to this as the principle that one can meaningfully speak about what it is to be "like" a conscious being — even if we can never actually have that experience beyond ourselves.

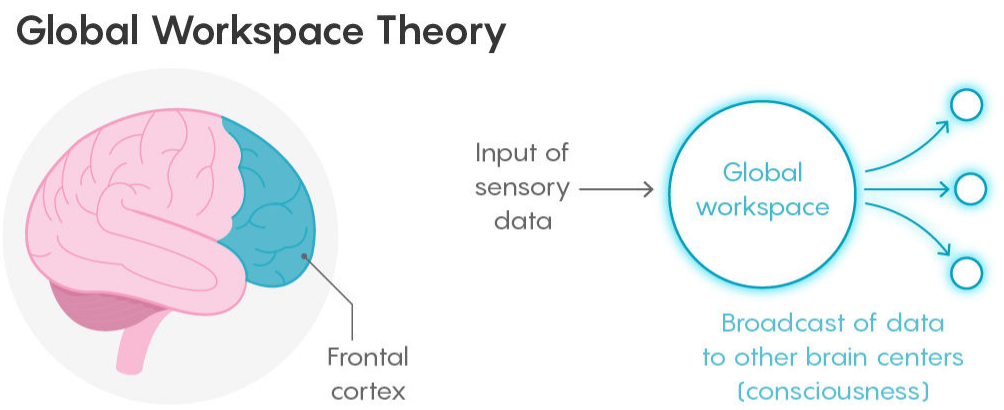
Plenty of cognition takes place outside the grasp of conscious awareness — in that sense, we respond to some cues and stimuli "unconsciously." A distinguishing feature of our minds, however, is that we can hold on to a piece of information, an idea or an intention as a motivation for subsequent decisions and behaviors. If we're hungry, we salivate as a reflex, but we might also choose to eat, go to the kitchen and get what we want from the cupboard.

Some researchers, such as the cognitive scientist Stanislas Dehaene of the Collège de

France in Paris, suggest that this conscious behavior arises when we hold a piece of information in a "global workspace" within the brain, where it can be broadcast to brain modules associated with specific tasks. This workspace, he says, imposes a kind of information bottleneck: Only when the first conscious notion slips away can another take its place. According to Dehaene, brain-imaging studies suggest this "conscious bottleneck" is a distributed network of neurons in the brain's prefrontal cortex.

This picture of consciousness is called global workspace theory (GWT). In this view, consciousness is created by the workspace itself — and so it should be a feature of any information-processing system capable of broadcasting information to other processing centers. It makes consciousness a kind of

computation for motivating and guiding actions. "Once you have information and the information is made broadly available, in that act consciousness occurs," said Christof Koch, chief scientist and president of the Allen Institute for Brain Science in Seattle.



According to one theory, consciousness is a form of information processing. It occurs when sensory data for an experience go to a "global workspace" and are distributed to other centers. The architecture for this process in the brain may be in the frontal cortex .

Lucy Reading-Ikkanda/Quanta Magazine

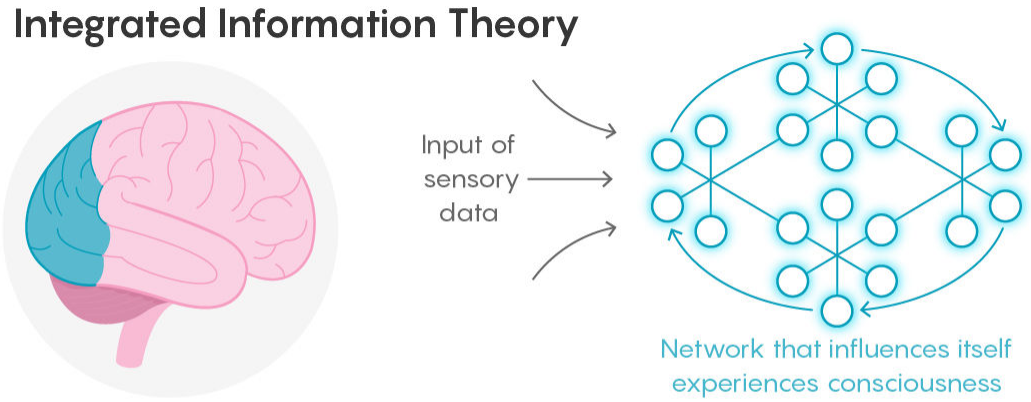
But to Koch, the argument that all of cognition, including consciousness, is merely a form of

computation "embodies the dominant myth of our age: that it's just an algorithm, and so is just a clever hack away." According to this view, he said, "very soon we'll have clever machines that model most of the features that the human brain has and thereby will be conscious."

He has been developing a competing theory in collaboration with its originator, the neuroscientist Giulio Tononi of the University of Wisconsin-Madison. They say that consciousness is not something that arises while turning inputs into outputs but rather an intrinsic property of the right kind of cognitive network, one that has specific features in its architecture. Tononi christened this view integrated information theory (IIT).

In contrast to GWT, which starts by asking

what the brain does to create the conscious experience, IIT begins instead with the experience. "To be conscious is to have an experience," Tononi said. It doesn't have to be an experience about anything, although it can be; dreams, or some "blank mind" states attained by meditation also count as conscious experiences. Tononi has sought to identify the essential features of these experiences: namely, that they are subjective (they exist only for the conscious entity), structured (their contents relate to one another: "the blue book is on the table"), specific (the book is blue, not red), unified (there is only one experience at a time) and definitive (there are bounds to what the experience contains). From these axioms, Tononi and Koch claim to have deduced the properties that a physical system must possess if it is to have some degree of consciousness.



The integrated information theory argues that consciousness is intrinsic to cognitive networks that exert a "causal power" on themselves. The back of the brain might have the right architecture for this capacity.

Lucy Reading-Ikkanda/Quanta Magazine

IIT does not portray consciousness as information processing but rather as the causal power of a system to "make a difference" to itself. Consciousness, Koch said, is "a system's ability to be acted upon by its own state in the past and to influence its own future. The more a system has cause-and-effect power, the more conscious it is."

This harks back to the famous "cogito, ergo sum" dictum of René Descartes in the 17th century. "The one thing, the only thing, that is [a] given is my experience," Koch said. "That's Descartes' central insight."

To Tononi and Koch, systems in which information is merely "fed forward" to convert inputs to outputs, as in digital computers, can only be "zombies," which might act as if they are conscious but cannot truly possess that property. Much of Silicon Valley may believe that computers will eventually become conscious, but to Koch, unless those machines have the right hardware for consciousness, they will just constitute a "deep fake."

"Digital computers can simulate consciousness, but the simulation has no causal power and is not actually conscious,"

Koch said. It's like simulating gravity in a video game: You don't actually produce gravity that way.

'Surrounded and Immersed' in Consciousness

One of the most striking features of IIT is that it makes consciousness a matter of degree. Any system with the required network architecture may have some of it. "No matter whether the organism or artifact hails from the ancient kingdom of Animalia or from its recent silicon offspring, no matter whether the thing has legs to walk, wings to fly, or wheels to roll with," Koch wrote in his 2012 book *Consciousness: Confessions of a Romantic Reductionist*. "If it has both differentiated and integrated states of information, it feels like something to be such a system."

This view arouses a lot of skepticism. The influential American philosopher of mind John Searle of the University of California, Berkeley derides the idea as a form of panpsychism: crudely, a belief that mind and awareness infuse the whole cosmos. In a withering critique of IIT, Searle has asserted that "the problem with panpsychism is not that it is false; it does not get up to the level of being false. It is strictly speaking meaningless because no clear notion has been given to the claim." Consciousness, he wrote, "cannot be spread over the universe like a thin veneer of jam" — it "comes in units and panpsychism cannot specify the units."

To Christof Koch, the chief scientist and president of the Allen Institute for Brain Science, the argument that consciousness is merely a form of computation "embodies the dominant myth of our age: that it's just an algorithm, and so is just a clever hack away." Unless artificial intelligences have the right features in their architecture, he says, they will never be more than "deep fakes."

—

Erik Dinnel

Koch, however, is perfectly happy to think that "we are surrounded and immersed" in consciousness. He believes "that consciousness is a fundamental, elementary property of living matter. It can't be derived from anything else."

But this doesn't mean it is spread equally everywhere. Koch and Tononi assert that, while consciousness can be an attribute of many things, a significant amount of it can exist only in particular kinds of things, notably human brains (indeed, in specific parts of human brains). And to turn IIT into a quantitative, testable theory, Koch and Tononi have formulated a criterion for what kinds of things those are.

To reflect how conscious an information-processing network is, Koch and Tononi define

a measure of "information integration," which they call Φ (the Greek letter phi). It represents the amount of "irreducible cause-effect structure": how much the network as a whole can influence itself. This depends on interconnectivity of feedback. If a network can be divided into smaller networks that don't exert causal power on one another, then it will have a correspondingly low value of Φ no matter how many processing nodes it has.

Equally, "any system whose functional connectivity and architecture yield a Φ value greater than zero has at least a trifle of [conscious] experience," Koch said. That includes the biochemical regulatory networks found in every living cell, and also electronic circuits that have the right feedback architecture. Since atoms can influence other atoms, "even simple matter has a modicum of

Φ." But systems that have enough Φ to "know" of their existence, as we do, are rare (although the theory anticipates that higher animals will also have a degree of that experience).

Because of this effort to make IIT quantitative and testable, Aaronson puts it "in something like the top 2 percent of all mathematical theories of consciousness ever proposed." He believes that the theory is flawed — but contrary to Searle, he says that "almost all competing theories of consciousness have been so vague, fluffy and malleable that they can only aspire to wrongness."

Seeking Neural Correlates

Koch would concur with that. "Everybody seems to have a pet theory of consciousness, but few of them are quantitative or predictive,"

he said. He believes that both GWT and IIT are testable. "Logically speaking, they could be wrong, or both could capture certain aspects of reality." How, though, do we test them?

Enter the Templeton World Charity Foundation, which has assigned $20 million to the task of testing theories of consciousness to destruction. It is starting with what Potgieter calls a "structured adversarial collaboration" involving IIT and GWT because they are able to make testable and contrasting predictions. The plan is for the proponents of the two theories to agree in advance to an experimental protocol that ought to distinguish whether either or both of the theories are wrong. "The condition was that the leaders of the theories would sign off on this protocol, in the sense of acknowledging that the predictions accurately represent the theory," Potgieter said. (He

credited the willingness of Dehaene and Tononi "to put themselves on the line" as one of the considerations that led to the choice of GWT and IIT as the first theories on the block.)

Dawid Potgieter of the Templeton World Charity Foundation is coordinating the proposed experimental program for testing pairs of consciousness theories head to head.

Templeton World Charity Foundation

The collaboration will get a top journal to commit to publishing the outcome of the experiments, come what may. The study will also include replication experiments. "This is basically open science," Potgieter said. "If we can use the best practices in open science to demonstrate progress in an area where no one has done very much, it could show that it's a useful approach."

He says that the researchers now have a final experimental design to test incompatible

predictions of GWT and IIT head-to-head. The details are yet to be disclosed, but they will deploy a battery of brain-monitoring techniques, such as functional magnetic resonance imaging (fMRI), electrocorticography and magnetoencephalography. The experiment seems to be "the first time ever that such an audacious, adversarial collaboration has been undertaken and formalized within the field of neuroscience," Potgieter added. He hopes that if the project is approved, the experimental work will be able to start after the summer and run for about three years, involving 10-12 labs.

What differences between the theories will the experiments test? One is in the location of consciousness in the brain. According to GWT, the "neural correlates of consciousness" —

the patterns of neuron activity that reflect the conscious state — should show up in parts of the brain that include the parietal and frontal lobes of the cortex. The parietal lobe processes sensory data such as touch and spatial sense. The frontal lobe is associated with cognitive processing for "higher" functions such as memory, problem solving, decision-making and emotional expression.

But people who have had a large fraction of the frontal lobe removed — as used to happen in neurosurgical treatments of epilepsy — can seem remarkably normal, Koch says. According to IIT, the seat of consciousness is instead likely to be in the sensory representation in the back of the brain, where the neural wiring seems to have the right character. "I'm willing to bet that, by and large, the back is wired in the right way to have

high Φ, and much of the front is not," Tononi said.

We can compare the locations of brain activity in people who are conscious or have been rendered unconscious by anesthesia, he says. If such tests were able to show that the back of the brain indeed had high Φ but was not associated with consciousness, he admits that "IIT would be very much in trouble."

A recent fMRI study of brain activity in volunteers who were either conscious or under general anesthesia, conducted by a group that included Dehaene, showed distinct patterns corresponding to the two states. During periods of unconsciousness, brain activity persisted only among regions with direct anatomical connections, whereas during conscious activity, complex long-distance

interactions did not seem constrained by the brain's "wiring."

However, one of the authors of the study, the physicist-turned-neuroscientist Enzo Tagliazucchi of the University of Buenos Aires and the Pitié-Salpêtrière Hospital in Paris, stresses that the findings don't yet clearly support any particular theory of consciousness. "It would be premature to frame our work within one theory or the other," he said. "It doesn't tip any balance, nor it is intended to do so."

Another prediction of GWT is that a characteristic electrical signal in the brain, arising about 300-400 milliseconds after a stimulus, should correspond to the "broadcasting" of the information that makes us consciously aware of it. Thereafter the

signal quickly subsides. In IIT, the neural correlate of a conscious experience is instead predicted to persist continuously while the experience does. Tests of this distinction, Koch says, could involve volunteers looking at some stimulus like a scene on a screen for several seconds and seeing whether the neural correlate of the experience persists as long as it remains in the consciousness.

"To be conscious is to have an experience," according to the neuroscientist Giulio Tononi of the University of Wisconsin-Madison. For that reason, the integrated information theory model of consciousness that he originated begins by asking what features a system must have to produce an experience like that of consciousness.

John Maniaci/UW Health

Not everyone is optimistic that it will be possible to find rigorous, definitive ways of testing and adjudicating these two theories. "The current project is an attempt in good faith in this direction," said Francis Fallon, a philosopher of mind at St. John's University in Queens, New York, who is involved in the Templeton project. But he noted that because both theories have already been shaped by

existing empirical evidence, it would be surprising to find new data with which either seems fundamentally inconsistent.

Hakwan Lau, a psychologist who studies behavioral neuroscience at the University of California at Los Angeles, is not convinced IIT is even a truly scientific theory. "IIT is based on armchair theorizing," he said. He thinks that what IIT advocates regard as the likely locus of consciousness doesn't necessarily follow from the theory but is just their subjective view. "To make empirical predictions [of the theory] testable by current methods," he said, "many additional assumptions and approximations need to be made."

To him, Lau says, IIT and GWT are "so different that I don't know how to begin to

compare them." In contrast, Tagliazucchi thinks it possible that the two are essentially the same theory, but "developed from third- and first-person viewpoints."

The cognitive scientist Anil Seth of the University of Sussex in the U.K. shares reservations about whether the Templeton project might prove premature. A "definitive rebuttal or validation" seems unlikely, he said, because the theories "make too many different assumptions, have different relations to testability and may even be trying to explain different things. GWT seems mostly about function and cognitive access, while IIT is a theory based primarily on phenomenology, not function, and is difficult to test."

Tononi and his collaborators would counter that they have been developing experimental

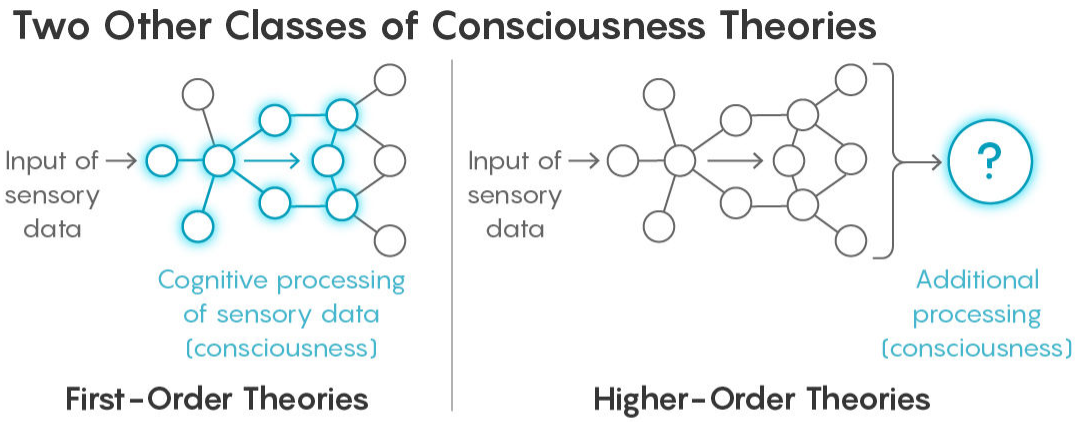
tests of IIT for many years — work that has led to the development of a crude but effective tool for evaluating consciousness in brain-damaged patients. Yet even Tononi agrees that, because both theories are still under construction and remain so "far apart," it might be too much to expect a definite outcome. "Their predictions aren't as precise as in physics," he said.

Still, he argues that "in the interests of making progress, you have to start with what you've got." Besides, the exercise "forces the theories to say something specific." Regardless of the outcome, Tononi feels sure that the tests will teach us something new and useful about the brain.

Other Contending Theories

No one imagines that eliminating GWT or IIT would solve the mystery of what consciousness is. For one thing, there are other serious theories, too.

Among them, two common classes are called first-order and higher-order theories (HOTs). "A first-order theory says that there's nothing more to the mind than the basic cognitive processing of sensory information," according to Lau. What brings some of that sensory information into consciousness, first-order theorists say, is something unidentified but intrinsic to how it's represented in the brain — for example, the dynamics of the interactions among elements in its neural network.



First-order theories maintain that consciousness is simply a product of the cognitive processing of sensory information. Higher-order theories posit that consciousness involves something done to build on that cognitive representation of the sensory experience.

Lucy Reading-Ikkanda/Quanta Magazine

In contrast, he said, "higher-order theorists say that the mind does something with the representation, over and above the cognition itself, to produce consciousness." In a HOT, a conscious experience is not merely a record of the perceptions involved but involves some additional mechanism that draws on that representation. That higher-order state doesn't necessarily serve some function in

processing the information, as in GWT; it just is.

"Compared to other existing theories, HOT can more readily account for complex everyday experiences, such as emotions and episodic memories," Lau and his colleagues, the philosopher Richard Brown of LaGuardia Community College and the neuroscientist Joseph LeDoux of New York University, wrote recently.

The Templeton World Charity Foundation has assigned further funds to test such ideas as it will GWT and IIT. "I hope to host about nine meetings over the next five years, to bring together two or more incompatible theories and try to hash it out between those theories," Potgieter said. He admits that "it might be that none of the current ideas is correct."

It may also turn out that no scientific experiment can be the sole and final arbiter of a question like this one. "Even if only neuroscientists adjudicated the question, the debate would be philosophical," Fallon said. "When interpretation gets this tricky, it makes sense to open the conversation to philosophers. Many of the neuroscientists in this field are already engaging in philosophy, some quite excellently."

Potgieter hopes that the adversarial approach will allow progress on other big questions — like understanding how consciousness arose in the first place, or how life itself did. "Wherever there is a big question with a bunch of different theories that are all strong but all siloed away from each other, we will try to make progress by breaking down the silos," he said.

"I think it is a wonderful initiative, and should be much more frequent in science," Tononi said. "It forces the proponents to focus and enter some common framework. I think we all stand to gain one way or another."