A Theory of Consciousness

Is complexity the secret to sentience, to a panpsychic view of consciousness?

BY CHRISTOF KOCH

DO YOU THINK that your newest acquisition, a Roomba robotic vacuum cleaner that traces out its unpredictable paths on your living room floor, is conscious? What about that bee that hovers above your marmalade-covered breakfast toast? Or the newborn who finally fell asleep after being suckled? Nobody except a dyed-in-the-wool nerd would

Is iRobot's Roomba robotic vacuum cleaner conscious? How about a bee, a comatose patient or a sleeping baby? Integrated information theory might tell us.

think of the first as being sentient; adherents of Jainism, India's oldest religion, believe that bees—and indeed all living creatures, small and large—are aware, whereas most everyone would accord the magical gift of consciousness to the baby.

The truth is that we really do not know which of these organisms is or is not conscious. We have strong feelings about the matter, molded by tradition, religion and law. But we have no objective, rational method, no step-by-step procedure, to determine whether a giv-

en organism has subjective states, has feelings.

The reason is that we lack a coherent framework for consciousness. Although consciousness is the *only* way we know about the world within and around us—shades of the famous Cartesian deduction cogito, ergo sum—there is no agreement about what it is, how it relates to highly organized matter or what its role in life is. This situation is scandalous! We have a detailed and very successful framework for matter and for energy but not for the mind-body problem. This dismal state of affairs might be about to change, however.

The universal lingua franca of our age is information. We are used to the idea that stock and bond prices, books, photographs, movies, music and our genetic makeup can all be turned into data streams of zeros and ones. These bits are the elemental atoms of information that are transmitted over an Ethernet cable or via wireless, that are stored, replayed, copied and assembled into gigantic repositories of knowledge. Information does not depend on the substrate. The same information can be represented as lines on paper, as electrical charges inside a PC's memory banks or as the strength of the synaptic connections among nerve cells.

Since the early days of computers, scholars have argued that the subjective, phenomenal states that make up the life of the mind are intimately linked to the information expressed at that time by the brain. Yet they have lacked the tools to turn this hunch into a concrete and predictive theory. Enter psychiatrist and neuroscientist Giulio Tononi of the Uni-

versity of Wisconsin–Madison. Tononi has developed and refined what he calls the integrated information theory (IIT) of consciousness.

An Integrated Theory

IIT is based on two axiomatic pillars.

First, conscious states are highly differentiated; they are informationally very rich. You can be conscious of an uncountable number of things: you can watch your son's piano recital, for instance; you can see the flowers in the garden outside or the Gauguin painting on the wall. Think of all the frames from all the movies you have ever seen or that have ever been filmed or that will be filmed! Each frame, each view, is a specific conscious percept.

Second, this information is highly integrated. No matter how hard you try, you cannot force yourself to see the world in black-and-white, nor can you see only the left half of your field of view and not the right. When you're looking

(The truth is that we **really do not know** which of these organisms is or is not conscious.)
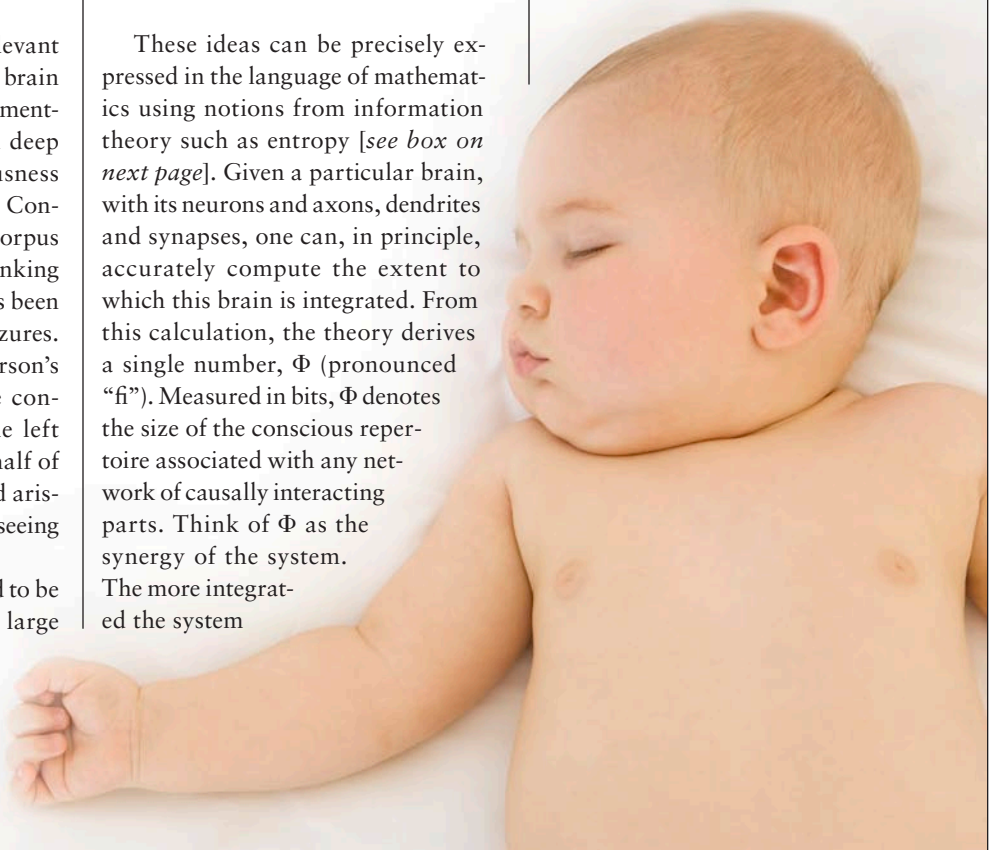
at your friend's face, you can't fail to also notice if she is crying. Whatever information you are conscious of is wholly and completely presented to your mind; it cannot be subdivided. Underlying this unity of consciousness is a multitude of

the family photographs on my Macintosh are not linked to one another. The computer does not know that the girl in those pictures is my daughter as she matures from a toddler to a lanky teenager and then a graceful adult. To my Mac, all information is equally meaningless, just a vast, random tapestry of zeros and ones.

Yet I derive meaning from these images because my memories are heavily cross-linked. And the more interconnected, the more meaningful they become. Indeed, Tononi's IIT postulates that the amount of integrated information that an entity possesses corresponds to its level of consciousness.

is, the more synergy it has, the more conscious it is. If individual brain regions are too isolated from one another or are interconnected at random, Φ will be low. If the organism has many neurons and is richly endowed with specific connections, Φ will be high—capturing the quantity of consciousness but not the quality of any one conscious experience. (That value is generated by the informational geometry that is associated with Φ but won't be discussed here.)

Explaining Brain Facts

The theory can account for a number of puzzling observations. The cerebellum, the "little brain" at the back of the brain that contains more neurons than the convoluted cerebral cortex that crowns the organ, has a regular, crystallinelike wiring arrangement. Thus, its circuit complexity as measured by Φ is

> (To be conscious, you must be a single, integrated entity with **a large repertoire** of highly differentiated states.)

causal interactions among the relevant parts of your brain. If areas of the brain start to disconnect or become fragmented and balkanized, as occurs in deep sleep or in anesthesia, consciousness fades and might cease altogether. Consider split-brain patients, whose corpus callosum—the 200 million wires linking the two cortical hemispheres—has been cut to alleviate severe epileptic seizures. The surgery literally splits the person's consciousness in two, with one conscious mind associated with the left hemisphere and seeing the right half of the visual field and the other mind arising from the right hemisphere and seeing the left half of the visual field.

To be conscious, then, you need to be a single, integrated entity with a large repertoire of highly differentiated states. Although the 60-gigabyte hard disk on my MacBook exceeds in capacity my lifetime of memories, that information is not integrated. For example,

These ideas can be precisely expressed in the language of mathematics using notions from information theory such as entropy [*see box on next page*]. Given a particular brain, with its neurons and axons, dendrites and synapses, one can, in principle, accurately compute the extent to which this brain is integrated. From this calculation, the theory derives a single number, Φ (pronounced "fi"). Measured in bits, Φ denotes the size of the conscious repertoire associated with any network of causally interacting parts. Think of Φ as the synergy of the system. The more integrated the system

(If you **lose your cerebellum,** you will never be a dancer or pianist, but your consciousness will be unimpaired.)

low as compared with that of the cerebral cortex. Indeed, if you lose your cerebellum you will never be a rock climber, pianist or ballet dancer, but your consciousness will not be impaired. The cortex and its gateway, the thalamus—the quail egg–shaped structure in the center of the brain—on the other hand, are essential for consciousness, providing it with its elaborate content. Its circuitry conjoins functional specialization with functional integration thanks to extensive reciprocal connections linking distinct cortical regions and the cortex with the thalamus. This cortico-thalamic complex is well suited to behave as a single dynamic entity endowed with a large number of discriminable states. Lose one chunk of a particular cortical area, and you might be unable to perceive motion. If a different area were lesioned, you would be blind to faces (yet could see the eyes, hair, mouth and ears).
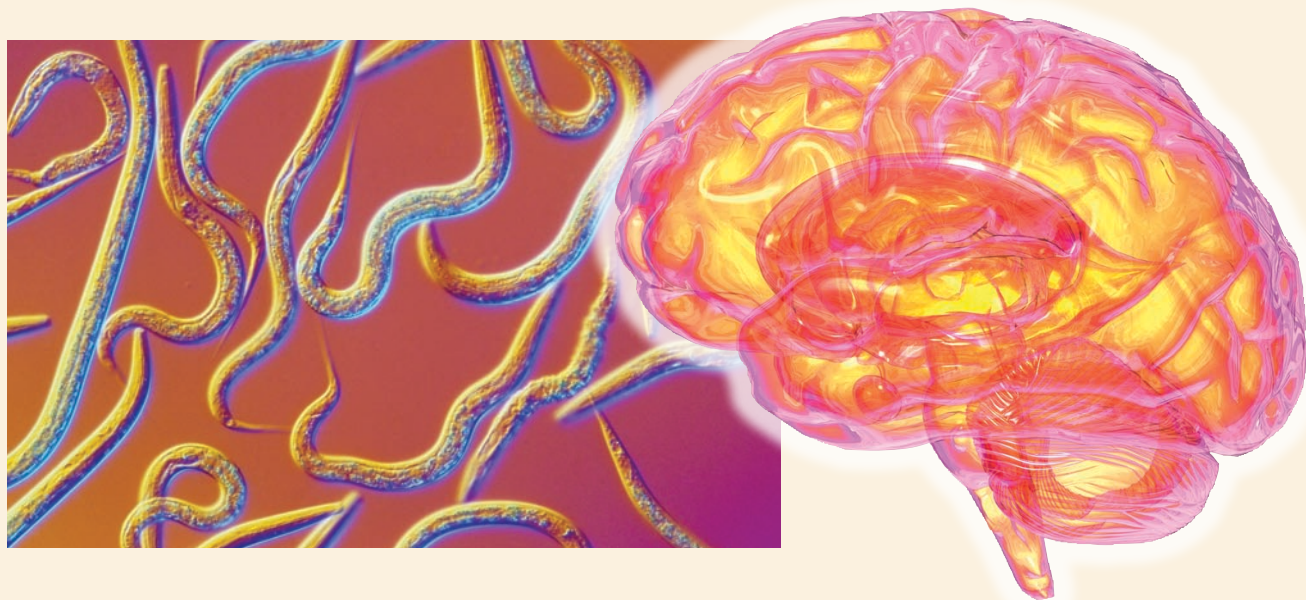
When people are woken from deep sleep, they typically recall experiencing nothing or, at best, only some vague bodily feeling; this experience contrasts with the highly emotional narratives our brains weave during rapid-eye-movement (REM) sleep. What is paradoxical is that the average firing activity of individual nerve cells does not differ that much in deep sleep and quiet wakefulness. At the whole system level, though, electroencephalographic electrodes on the skull pick up slow, large and highly synchronized waves during deep sleep. Because these waves are quite regular, they will disrupt the transfer of specific information among brain cells.

Every day, in tens of thousands of surgical operations, patients' consciousness is quickly, safely and transiently turned off and on again with the help of

How to Calculate Consciousness in All Creatures

Integrated information theory uses mathematics (*bottom*) to quantify the amount of integrated information an entity possesses—and thus its level of consciousness. The challenge: we cannot yet calculate the state of awareness for even the simple roundworm (*left*) with current computers, let alone deal with the complexity of the human brain.

$$ei(x_1; P) = -\sum_{i=1}^{k} \sum_{\mu_0^{(i)}} p(\mu_0^{(i)}|x_1) \log p(\mu_0^{(i)}|\mu_1^{(i)}) - H(X_0|x_1)$$

$$\Phi(x_1) = \min_P \frac{ei(x_1; P)}{v_P}$$

various anesthetic agents. There is no single mechanism common to all. The most consistent regional finding is that anesthetics reduce thalamic activity and deactivate mesial (middle) and parietal cortical regions. Twenty years of electrical recording in anesthetized laboratory animals provided ample evidence that many cortical cells, particularly in primary sensory cortical regions, continue to respond selectively during anesthesia. What appears to be disrupted is large-scale functional integration in the corticothalamic complex.

IIT explains why consciousness requires neither sensory input nor behavioral output, as happens every night during REM sleep, in which a central paralysis prevents the sleeper from acting out her dreams. All that matters for consciousness is the functional relation among the nerve cells that make up the corticothalamic complex. Within this integrated dynamic entity can be found the dream of the lotus eater, the mindfulness of the meditating monk, the agony of the

amount of synergy, of Φ. In this sense, IIT is a scientific version of panpsychism, the ancient and widespread belief that all matter, all things, animate or not, are conscious to some extent. Of course, IIT does not downplay the vast gulf that separates the Φ of the common roundworm *Caenorhabditis elegans* with its 302 nerve cells and the Φ associated with the 20 billion cortical neurons in a human brain.

The theory does not discriminate between squishy brains inside skulls and silicon circuits encased in titanium. Provided that the causal relations among the transistors and memory elements are complex enough, computers or the billions of personal computers on the Internet will have nonzero Φ. The size of Φ could even end up being a yardstick for the intelligence of a machine.

Future Challenges

IIT is in its infancy and lacks the graces of a fully developed theory. A major question that it so far leaves un-

charge is a fundamental feature of the universe without a function, consciousness might also lack any specific evolutionary role. It just is.

A second stumbling block with IIT is that Φ is exceedingly difficult to compute even for very small systems. To accurately evaluate Φ for the roundworm is utterly unfeasible, even if using all of Google's more than 100,000 computers. Can we find other algorithms to more easily compute Φ?

A third issue to understand is why so much brain processing and so many of our daily behaviors are unconscious. Do the neural networks that mediate these unconscious, zombielike behaviors have lower Φ than the ones that give rise to consciousness?

Tononi's integrated information theory of consciousness could be completely wrong. But it challenges us to think deeply about the mind-body problem in a novel, rigorous, and mathematically and empirically minded manner. And that is a great boon to this endeavor.

> (The theory **does not discriminate** between squishy brains inside skulls and silicon circuits encased in titanium.)

cancer patient and the Arcadian visions of your lost childhood home. Paraphrasing Oscar Wilde, I would say it is the causal interactions within the dynamic core that make the poppy red, the apple odorous and the skylark sing.

Consciousness Is Universal

One unavoidable consequence of IIT is that all systems that are sufficiently integrated and differentiated will have some minimal consciousness associated with them: not only our beloved dogs and cats but also mice, squid, bees and worms.

Indeed, the theory is blind to synapses and to all-or-none pulses of nervous systems. At least in principle, the incredibly complex molecular interactions within a single cell have nonzero Φ. In the limit, a single hydrogen ion, a proton made up of three quarks, will have a tiny

answered is, Why should natural selection evolve creatures with high Φ? What benefit for the survival of the organism flows from consciousness? One answer that I hope for is that intelligence, the ability to assess situations never previously encountered and to rapidly come to an appropriate response, requires integrated information. Another possible answer, though, could be that high-Φ circuits do not have any special status in terms of their survival. Just as electrical

If Tononi's equation for Φ proves to plumb the hitherto ineffable—consciousness itself—it would validate the ancient Pythagorean belief that "number is the ruler of forms and ideas and the cause of gods and demons." **M**

CHRISTOF KOCH is Lois and Victor Troendle Professor of Cognitive and Behavioral Biology at the California Institute of Technology and serves on *Scientific American Mind*'s board of advisers.

(Further Reading)

◆ **Can Machines Be Conscious?** Christof Koch and Giulio Tononi in *IEEE Spectrum,* Vol. 45, No. 6, pages 55–59; June 2008.
◆ **Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework.** David Balduzzi and Giulio Tononi in *PLoS Computational Biology,* Vol. 4, No. 6; June 2008.
◆ **Consciousness as Integrated Information: A Provisional Manifesto.** Giulio Tononi in *Biological Bulletin,* Vol. 215, No. 3, pages 216–242; December 2008.