A Bit of Theory: Consciousness as Integrated Information

This is part of IEEE Spectrum's SPECIAL REPORT: THE SINGULARITY

PHOTO: Giulio Tononi

If the level of consciousness has to do with how much integrated information a conscious entity generates, it is essential that we determine to what extent different neural architectures can generate integrated information. The integrated-information theory of consciousness, or IIT, is an attempt to do so, and to approach consciousness from first principles.

IIT introduces a measure of integrated information, represented by the symbol Î¿ and given in bits, that quantifies the reduction of uncertainty (that is, the information generated when a system enters a particular state through causal interactions among its parts) This measure is above and beyond the information generated independently within the parts themselves. The parts should be chosen in such a way that they can account for as much nonintegrated (independent) information as possible.

If a system has a positive value of Î¿ (and it is not included within a larger subset having higher Î¿), it is called a complex. When a complex enters a particular state of its repertoire, it generates an amount of integrated information corresponding to Î¿. Thus, a simple photodiode that can detect the presence or absence of light is a complex with Î¿=1 bit. The sensor chip of a digital camera, on the other hand, would not form a complex: as such it would have Î¿=0 bits, as each photodiode

does its job independently of the others. In principle, it can be decomposed into individual photodiodes, each with Î¿=1 bit.

Within the awake human brain, on the other hand, there must be some complex whose Î¿ value is on average very high, corresponding to our large repertoire of conscious states that are experienced as a whole. Because integrated information can be generated only within a complex and not outside its boundaries, it follows that consciousness is necessarily subjective, private, and related to a single point of view or perspective.

Given the vast number of ways even a simple information-processing system can be decomposed, measuring Î¿ can be done only for exceedingly simple systems. At this point there's no Î¿-meter to determine how much consciousness resides in an iPhone, a dog, a newborn baby, a comatose patient, or the brain of Ray Kurzweil. Also, the value of Î¿ depends on both spatial and temporal scales that determine what counts as elements and states of a system. We speculate that the relevant spatial and temporal scales are those that jointly maximize Î¿, but properly testing this idea is even less feasible. Right now, the only practical way to test the validity of the integrated-information approach is to consider some predictions of the theory and compare these against neurobiological data.

With the aid of computer simulations, one can try out different networks and see which architectures yield high values of Î¿, everything else (the number of elements and connections, for instance) being equal. Such simulations invariably indicate that high Î¿ requires networks that combine functional specialization with functional integration so that each element has a unique function within the network and there are many pathways for interactions among the elements. In very rough terms, this kind of architecture describes the mammalian thalamocortical system: different parts of the cerebral cortex are specialized for different functions, yet a vast network of connections allows these parts to interact profusely. And indeed, the thalamocortical system is precisely the part of the brain that cannot be severely impaired without the loss of consciousness.

Conversely, Î¿ is low for systems made up of small, quasi-independent modules. This suggests that parts of the brain that are organized in an extremely modular

manner, where modules hardly interact, should not contribute directly to consciousness. The cerebellum at the back of the skull offers a dramatic demonstration. While smaller than the cortex, the cerebellum has more neurons (some 50 billion), lots of connections, outputs that exert fine control on behavior, and a massive endowment of neurotransmitters and neuromodulators. And yet, if the cerebellum is damaged due to stroke, trauma, or some other calamity, consciousness is largely unaffected. Interestingly, the synaptic organization of the cerebellum is such that individual sectors are activated independently of one another, with little interaction between distant ones. Thus, although the cerebellum is a powerful computer, it is the wrong machine for consciousness, being far too modular to generate much integrated information

Computer simulations also indicate that parallel input or output pathways can be attached to a complex of high Î¿ without becoming part of it. This may explain, for example, why the retina, which is connected to the visual cortex by multiple parallel pathways, does not directly contribute to visual consciousness.

Simulations show that a complex of high Î¿ can be augmented by attaching local circuits to some of its elements—circuits that take local inputs, process them, and then return them locally—and yet the attached circuits may remain outside of the high Î¿ complex. In the brain, it appears that many computations that remain unconscious are carried out by cortical and subcortical circuits that appear to be informationally insulated. For example, we remain unaware of the elaborate processes that enable us to see in depth, recognize an object, or parse a sentence. Instead, we are only aware of the results of such computations—we see meaningful objects laid out in space and separated from the background and hear meaningful sentences.

The situation is similar on the executive side of consciousness. Our vague intentions are miraculously translated into the right words, strung together to form a syntactically correct sentence that conveys what we meant to say. Yet again, we are not conscious of the underlying processing, much of which takes place in the cortex.

These and other examples illustrate that consciousness is associated with neural

architectures that form a complex of high Îż: a single entity having a large repertoire of states. Though it is too early to know whether these simple examples scale up, it seems that an excellent way to achieve high values of Îż is to build a network that is both highly specialized and highly integrated--one that simultaneously achieves the benefits of efficient categorization and association. Such a network could be augmented by attaching many local subroutines that would remain informationally insulated. By contrast, completely modular architectures, or homogeneous ones, are definitely not the way to go.

According to this theory, Îż, and therefore conscious experience, is graded. It is not an all-or-none property that only sufficiently complex systems possess. Any physical system with some capacity for integrated information would have some degree of conscious experience, irrespective of the stuff of which it is made and independent of its ability to report it. The question, of course, is at what level of Îż consciousness becomes significant, both practically and ethically.

Here, too, we have only our first-person evidence to fall back on. When we are in a dreamless sleep or anesthetized, we ”lose” consciousness, even though our brains are not electrically inert. However, experiments using transcranial magnetic stimulation and EEG recordings of cortical activity have shown that the brain's ability to integrate information breaks down during dreamless sleep, which is consistent with the theory. Practically speaking, we can then think of the Îż value associated with dreamless sleep or general anesthesia as a threshold for consciousness. Anything with a smaller Îż is no more conscious than you or I during such a dreamless state. If we had to exist permanently at Îż values as low or lower than that--think of Terri Schiavo, who lived for 15 years in a persistent vegetative state--we would presumably not care, as it would feel like nothing at all.

For more articles, videos, and special features, go to The Singularity Special Report