Consciousness and the brain: theoretical aspects

Giulio Tononi

Consciousness is everything we experience, from a city scene teeming with shapes, colors, and noises to darkness and silence. It has been defined as what abandons us every night when we fall into dreamless sleep and returns the next morning when we wake up (**Tononi and Edelman 1998**). We know that consciousness depends on the functioning of the brain; yet many philosophers have claimed that understanding how the brain produces subjective experience may lie forever beyond the realm of scientific explanation. Scientists often find it safer to dispense with explanations and to explore instead which parts of the brain are important for consciousness and which aspects of neural activity accompany it. For example, clinical observations and classic experiments have shown that the level of consciousness is associated with an activated electroencephalogram. Brain lesions have demonstrated that different brain areas are important for specific contents of consciousness. New experimental tools are refining our appreciation of when and where changes in neural activity are correlated with changes in conscious experience.

Important as it is, however, gathering data on the neural correlates of consciousness is no substitute for a principled explanation. Otherwise, how can we establish whether newborn babies or animals are conscious and to what extent, or whether a conscious artifact can be constructed with nonneural ingredients? And how can we determine, in the absence of verbal reports, whether a person with akinetic mutism—awake with eyes open, but mute, immobile, and unresponsive—is conscious or not, or how much consciousness there is during sleepwalking or psychomotor seizures? A genuine understanding of consciousness requires, in addition to empirical studies, a theoretical analysis.

1. The two problems of consciousness

Two main problems need to be addressed. The first problem has to do with the

necessary and sufficient conditions that determine whether consciousness is present or not. For example, why are changes of neural activity in thalamocortical regions so important for conscious experience, whereas changes in activity in cerebellar circuits are not, given that the number of neurons in the two structures is comparable? Why is consciousness strikingly reduced during deep slow-wave sleep, given that the level of neuronal firing is similar to that of wakefulness?

The second problem has to do with the necessary and sufficient conditions that determine the specific nature of consciousness. For example, what makes the activity of specific cortical areas contribute specific dimensions of conscious experience—auditory cortex to sound, visual cortex to shapes and colors? What aspect of neural organization is responsible for the fact that shapes look the way they do, and different from the way colors appear or pains feel? Solving the first problem means that we would know whether a physical system can generate consciousness and to what extent—the *level* of consciousness. Solving the second problem means that we would know what kind of consciousness is generated—the *content* of consciousness.

2. Consciousness and the brain: facts and enigmas

Let us briefly consider the first problem by examining some facts about consciousness and the brain and some of the baffling questions they pose:

(1) Consciousness is produced by certain parts of the brain and not, or much less, by others. For example, we know that certain portions of the thalamocortical system are essential for consciousness. Other regions of the brain, such as the cerebellum, are not and can be stimulated or lesioned without giving rise to changes in conscious experience. Yet the cerebellum has as many neurons and is every bit as complicated as the thalamocortical system. Why is consciousness associated with some but not with other neural structures?

(2) Consciousness is associated with a distributed neural system: there is no single area where it all comes together. For example, different areas of the thalamocortical system each contribute different dimensions to conscious experience, but no single area is solely responsible for consciousness.

(3) Consciousness can be split if the thalamocortical system is split. Studies of split brain patients, whose corpus callosum was sectioned for therapeutic reasons,

show that each hemisphere has its own, private conscious experience.

(4) Within the thalamocortical system, the activity of certain neurons correlates well with conscious experience; that of others does not. For example, during binocular rivalry, the activity of certain visual cortical neurons follows what the subject consciously perceives, while that of other neurons follows the stimulus, whether the subject is perceiving it or not. What determines whether the firing of neurons within the thalamocortical system contributes directly to consciousness or not?

(5) Many neural processes in the thalamocortical system can influence conscious experience yet do not seem to contribute directly to it. For example, what we see and hear depends on elaborate computational processes in the cerebral cortex that are responsible for object recognition, depth perception, and language parsing; yet such processes remain largely unconscious.

(6) Neural activity in sensory afferents to the thalamocortical system usually determines what we experience at any given time. However, such neural activity does not appear to contribute directly to conscious experience. For example, although retinal cells can tell light from dark and convey that information to visual cortex, their rapidly shifting firing patterns do not correspond well with what we perceive. Moreover, a person who becomes retinally blind as an adult continues to have vivid visual images and dreams. Why is it that the activity of retinal cells *per se* does not contribute directly to conscious experience, but only indirectly through its action on thalamocortical circuits?

(7) Neural processes occurring in brain regions whose inputs and outputs are closely linked to the thalamocortical system, such as the basal ganglia, are important in the production and sequencing of action, thought, and language. Yet such processes do not seem to contribute directly to conscious experience. Why is this so?

(8) When first learning a new task, we are conscious of every detail of the task; we are slow, make mistakes, and must make an effort. When we have learned the task well, we become much better, faster, and effortless, but also less aware of it. During learning, many cortical areas are involved, whereas well-learned tasks may employ different neural circuits. Why does automatic performance become less conscious?

(9) The mode of neuronal firing in the thalamocortical system is important for consciousness. For example, cortical neurons fire almost as much during deep slow-wave sleep as during wakefulness, but the level of consciousness is much reduced in the former condition. Similarly, in absence seizures, neural firing is high and synchronous; yet consciousness is seemingly lost. Why is this the case?

(10) The firing of the same cortical neurons may correlate with consciousness at certain times, but not at others. For example, multiunit recordings in the primary visual cortex of monkeys show that, after a stimulus is presented, the firing rate of many neurons increases regardless of whether the animal reports seeing a figure or not. After 80 to 100 milliseconds, however, their discharge accurately predicts the conscious detection of the figure. What determines when the firing of the same cortical neurons contributes to conscious experience and when it does not?

Many more facts and puzzles could be added to this list. This state of affairs is not unlike the one biologists faced when, knowing a great deal about similarities and differences between species, fossil remains, and breeding practices, they still lacked a theory of how evolution might occur. What was needed, then as now, were not just more facts, but a theoretical framework that could make sense of them. Unfortunately, theoretical approaches that try to provide a coherent explanation for some of the basic facts about consciousness and the brain are few and far between. Here, in order to offer a tentative but at least unified perspective on the issues that need to be addressed, we discuss a theoretical approach, according to which consciousness is related to the brain's ability to integrate information (**Tononi**, 2001, 2003). This theory comprises (a) an examination of phenomenology indicating that consciousness has to do with information integration, (b) a definition of what information integration is and how it can be measured, (c) an attempt at accounting for basic facts about consciousness and the brain, and (d) some corollaries and predictions.

3. Phenomenology: consciousness as integrated information   top

In this view, perhaps the most important thing to realize about consciousness is that, when one experiences a particular conscious state—for example, the one experienced when reading *this particular phrase* here and now—each of us is gaining access to an extraordinarily large amount of information. This information has nothing to do with how many letters or words we can take in at any one time, which is a very small number. Instead, the occurrence of a

particular conscious state is extraordinarily informative because of the very large number of alternative conscious states that it rules out. Just think of all possible written phrases you could read, multiply them by the number of possible fonts, ink colors, sizes in which you could read them; then think of the same phrases spoken aloud, or read and spoken, or think further of all other possible visual scenes you might experience, multiplied by all possible sounds you might hear at the same time, by all possible moods you might be in, and so on *ad libitum*.

The point is simply this: every time we experience a particular conscious state out of such a huge repertoire of possible conscious states, we gain access to a correspondingly large amount of information. This conclusion is in line with the classical definition of information as reduction of uncertainty among a number of alternatives (**Shannon and Weaver, 1949**). For example, tossing a fair coin and obtaining heads corresponds to $\log_2(2) = 1$ bit of information, because there are just two alternatives; throwing a fair die yields $\log_2(6) = 2.59$ bits of information, because there are six equally likely possibilities. Similarly, the information generated by the occurrence of a particular conscious state lies in the large number of different conscious states that *could potentially* have been experienced but were not. Although no attempt has been made to estimate the size of the repertoire of conscious states available to a human being, it is clear that such repertoire must be extraordinarily large, and so is the information yielded by entering a particular conscious state out of this repertoire. This point is so simple that its importance has been overlooked.

Another key aspect of this theory is that the information associated with the occurrence of a conscious state is information from the perspective of an *integrated* system. When each of us experiences a particular conscious state, that conscious state is experienced as an integrated whole—it cannot be subdivided into independent components, that is, components that are experienced independently. For example, the conscious experience of *this particular phrase* cannot be experienced as subdivided into, say, the conscious experience of how the words look independently of the conscious experience of how they sound in our mind. Similarly, we cannot experience visual shapes independently of their color, or perceive the left half of the visual field of view independently of the right half. Furthermore, we cannot simultaneously experience both aspects of an ambiguous figure, nor can we simultaneously think two different thoughts. If we could, this would be tantamount to having two separate "centers" of consciousness. Separate centers of consciousness exist, of course, but each is a

different person with a different brain (or a different hemisphere of a split brain). Indeed, if two people each experience a different conscious state in two separate rooms, the information to which each person gains access depends only on the repertoire of conscious states available to that person. A superordinate consciousness associated with the joint states of two different people considered together is an absurd notion, because the two people are causally independent and do not constitute an integrated system.

Finally, it is important to appreciate that every conscious experience takes some time. Studies of how a percept is progressively specified and stabilized—a process called *microgenesis*—indicate that it takes up to 100 to 200 milliseconds to develop a fully formed sensory experience and that the surfacing of a conscious thought may take even longer. Other evidence indicates that a single conscious moment cannot extend beyond 2 to 3 seconds. Although it is arguable whether conscious experience unfolds more like a series of discrete snapshots or a continuous flow, its time scale is certainly comprised between these lower and upper limits. Thus a phenomenological analysis indicates that consciousness has to do with the ability to integrate a large amount of information over a short period of time.

4. Measuring information integration

If consciousness is integrated information, then a physical system should be able to generate consciousness to the extent that it can rapidly enter any of a large number of available states (information); yet it cannot be decomposed into a collection of causally independent subsystems (integration). How can one identify such an integrated system, and how can one measure its repertoire of available states?

At first sight, it might seem that all one needs to do is choose a system, for example, the brain, and measure the repertoire of states that are available to it with their probability. One could then calculate the information associated with the occurrence of each brain state, just as one can measure the information associated with tossing a coin or a die, by using the entropy function, that is, the weighted sum of the logarithm of the probability of system states: $H = -\Sigma p(s) \log_2 p(s)$. Measuring the available repertoire would easily account for why a seemingly similar task can be performed unconsciously (or nearly so) by a simple device and consciously by a human being. For example, when a retinal cell, or even a photodiode—a simple semiconductor device that changes its

electrical resistance depending on the illumination—detects complete darkness, it generates a minimal amount of information, since it can only discriminate between darkness and light. When we consciously detect complete darkness, however, we perform a discrimination that is immensely more informative: we are not just ruling out light, but an extraordinary number of other possible states of affairs, including every possible frame of every possible movie, every possible sound, and every possible combination of them.

Measuring information this way, however, is insufficient, because it is completely insensitive to whether the information is available to an integrated system. To give a simple example, consider a collection of one million photodiodes constituting the sensor chip of a digital camera. From the perspective of an external observer, such a chip can certainly enter a very large number of different states, as could easily be demonstrated by presenting it with all possible input signals. However, because of the absence of any physical interaction among the photodiodes, the chip as such cannot integrate any information: the state of each element is causally independent of that of other elements. In other words, what we have is one million photodiodes with a repertoire of two states each, rather than a single integrated system with a repertoire of $2^{1,000,000}$ states. Thus to measure integrated information, it is essential to know whether a set of elements constitutes a causally integrated system or can be broken down into a number of independent or quasi-independent subsets among which no information can be integrated.

To see how one can achieve this goal, consider an extremely simplified neural system composed of a set of elements. Each element could represent, for instance, a group of locally interconnected neurons that share inputs and outputs, such as a cortical minicolumn. We could further assume that each element can go through discrete activity states, corresponding to different firing levels, each of which lasts for a few hundred milliseconds. Finally, for the present purposes, let us imagine that the system is disconnected from external inputs, just as the brain is disconnected from the environment when it is dreaming.

4.1. Effective information

Consider now a subset S of elements taken from the system. We want to measure the information generated when S enters a particular state out of its repertoire, but only to the extent that such information can be integrated (i.e., it can be the result of causal interactions within the system). To do so, we start by dividing S

into two complementary parts A and B, give maximum entropy to the outputs from A (i.e., substitute its elements with independent noise sources) and determine the entropy of the resulting responses of B. Specifically, we define the *effective information* between A and B as $EI(A \rightarrow B) = MI(A^{Hmax} \rightarrow B)$. Here MI $(A;B) = H(A) + H(B) - H(AB)$ stands for mutual information, the standard measure of the entropy or information shared between a source (A) and a target (B). Since A is substituted by independent noise sources, the entropy that B shares with A is due to causal effects of A on B. In neural terms, we try out all possible combinations of firing patterns as outputs from A and establish how differentiated is the repertoire of firing patterns they produce in B. Thus if the connections between A and B are strong and specialized, different outputs from A will produce different firing patterns in B, and $EI(A \rightarrow B)$ will be high. On the other hand, if the connections between A and B are such that different outputs from A produce scarce effects, or if the effect is always the same, then $EI(A \rightarrow B)$ will be low or zero. Note that, unlike measures of statistical dependence, effective information measures causal interactions and requires perturbing the outputs from A. Moreover, by enforcing independent noise sources in A, effective information measures all possible effects of A on B, not just those that are observed if the system were left to itself. Also, $EI(A \rightarrow B)$ and $EI(B \rightarrow A)$ will generally differ; that is, effective information is not symmetric. For a given bipartition of a subset, the sum of the effective information for both directions is indicated as $EI(A \leftrightharpoons B) = EI(A \rightarrow B) + EI(B \rightarrow A)$.

4.2. Information integration

Based on the notion of effective information for a bipartition, we can assess how much information can be integrated within a system of elements. To this end, it is important to realize that a subset S of elements cannot integrate any information if there is a way to partition S in two complementary parts A and B such that EI $(A \leftrightharpoons B) = 0$. In such a case we would obviously be dealing with two (or more) causally independent subsets, rather than with a single, integrated subset. More generally, to measure how much information can be integrated within a subset S, we should search for the bipartition(s) for which $EI(A \leftrightharpoons B)$ reaches a minimum. Since $EI(A \leftrightharpoons B)$ is necessarily bounded by the maximum entropy available to A or B, to be comparable over bipartitions, $\min\{EI(A \leftrightharpoons B)\}$ should be normalized by $\min\{H^{max}(A), H^{max}(B)\}$. Thus the *minimum information bipartition* of subset S, or MIB(S), is its bipartition for which the normalized effective information reaches a minimum, corresponding to $\min\{EI(A \leftrightharpoons B) \; / \; (\min\{H^{max}(A), H^{max}$ (B)\})\}$ for all A. The capacity for *information integration* of subset S, or $\Phi(S)$, is simply the value of $EI(A \leftrightharpoons B)$ for the minimum information bipartition: $\Phi(S) = EI$ (MIB(S)). The symbol Φ is meant to indicate that the information (the "I") can be

integrated within a single entity (the "O"). This quantity is also called ^{MIB}*complexity*, for minimum information bipartition complexity. In most cases, the bipartitions for which the normalized value of EI will be at a minimum, everything else being equal, will be bipartitions that cut the system in two halves, that is, midpartitions (**Tononi, 2001**).

4.3. Complexes

If Φ(S) is calculated for every possible subset S of a system, we can establish which subsets are actually capable of integrating information, and how much of it. Consider thus every possible subset S of m elements out of the n elements of a system, starting with subsets of two elements (m = 2) and ending with a subset corresponding to the entire system (m = n). A subset S with Φ >0 is called a *complex* if it is not included within a larger subset having higher Φ. For a given system, the complex with the maximum value of Φ(S) is called the *maximum* or *main complex*, where the maximum is taken over all combinations of m >1 out of n elements of the system. Note that a complex can be causally connected to elements that are not part of it. The elements of a complex that receive inputs from or provide outputs to other elements not part of that complex are called *ports in* and *ports out*, respectively. In summary, a system can be analyzed to identify its complexes—those subsets of elements that integrate information—and each complex will have an associated value of Φ—the amount of information it integrates.

5. Consciousness, information integration, and the brain

If consciousness has to do with the ability to integrate information, and if such ability can be measured as suggested above, it follows that a physical system will have consciousness to the extent that it constitutes a complex having high values of Φ. How do these concepts apply to the brain, and can they account, at least in principle, for some of the facts and puzzles listed above? Can they shed any light, for instance, on why the thalamocortical system is essential for consciousness whereas the cerebellum is not, or on why consciousness is reduced during slow-wave sleep?

Although performing an exhaustive analysis of complexes on the brain itself is not feasible at present, it is nevertheless possible to obtain some insights by considering computer simulations that capture basic aspects of brain organization (**Tononi et al., 1994**; **Sporns et al., 2000**). To be sure, the effective information between subsets of neurons is heavily constrained by biological factors affecting

maximum firing rates, firing duration, and postsynaptic excitability. Everything else being equal, however, the single most important determinant of a network ability to integrate information is its anatomic organization. Connection patterns can determine whether a system is highly modular (i.e., splits into small, nearly independent complexes) or whether it forms a single, large complex. Moreover, complexes having the same number of elements and connections may have widely different values of Φ depending on how the connections are arranged. It is worth considering a few representative instances.

5.1. Thalamocortical system

Using small networks of linear elements, it is possible to show that Φ is maximized by having each element develop a different connection pattern with the rest of the complex (functional specialization) while ensuring that, no matter how the system is divided in two, the two parts are well connected (functional integration). The thalamocortical system—the brain structure most directly associated with consciousness—appears to be an excellent candidate for the rapid integration of information. It is composed of a large number of widely distributed elements that are functionally specialized. At the same time, these elements are intricately interconnected and exhibit short- and long-range synchronization, a hallmark of functional integration (**Tononi and Edelman, 1998**). Indeed, computer simulations suggest that the connectivity of the visual thalamocortical system, characterized by many local groups of heavily connected neurons, forward connections that increase functional specialization, back connections that can distribute information widely, and a large number of entangled short loops, is well suited to generating high values of Φ. It is not yet clear to what extent the connectivity may differ in other cortical regions, or whether certain cortical layers and cell types may play a privileged role in integrating information among distant areas. However, it is clear that the organization of the thalamocortical system is well suited to forming a complex capable of integrating a large amount of information.

5.2. Cerebellum

This brain region contains probably more neurons and as many connections as the cerebral cortex, receives mapped inputs from the environment, and controls several outputs. However, in striking contrast to the thalamocortical system, lesions or ablations indicate that the direct contribution of the cerebellum to conscious experience is minimal. As suggested by computer simulations, the reason may be that the organization of synaptic connections within the

cerebellum is radically different from that of the thalamocortical system and is not well suited to the integration of information. Specifically, the organization of the connections is such that individual patches of cerebellar cortex tend to be activated independently of one another, with little interaction between distant patches. This suggests that cerebellar connections may not be organized so as to generate a large complex of high Φ, but rather very many small complexes each with a low value of Φ.

5.3. Split brain

When the corpus callosum is sectioned, consciousness is split. The level of consciousness of the dominant hemisphere and most of its contents are not altered severely after the operation. The nondominant hemisphere also appears to be conscious, although it loses some important abilities. Some information, for example emotional arousal, seems to be shared across the hemispheres, probably thanks to subcortical common inputs. As illustrated by simple computer models, a "callosal" cut produces, out of a large complex corresponding to the connected thalamocortical system, two separate complexes. However, because there is great redundancy between the two hemispheres, their Φ value is not greatly reduced compared with when they formed a single complex. The analysis of complexes also identifies a complex corresponding to both hemispheres and their subcortical common inputs, although with much lower Φ values. That is, there is a sense in which the two hemispheres still form an integrated entity, but the information they share is minimal.

5.4. Cortical input and output systems

Why is it that retinal cells do not appear to contribute directly to visual consciousness, given that they evidently influence it by providing powerful signals to the thalamocortical system? According to computer simulations, input circuits providing powerful inputs to a complex add nothing to the information it can integrate if their effects are entirely accounted for by ports in, that is, by the elements of the complex to which they connect. Similar considerations apply to output circuits triggered by ports out.

5.5. Basal ganglia and cortico-subcortical loops

The basal ganglia are large neural structures that contain many circuits arranged in parallel, some implicated in motor and oculomotor control, others, such as the dorsolateral prefrontal circuit, in cognitive functions, and others, such as the lateral orbitofrontal and anterior cingulate circuits, in social behavior, motivation,

and emotion. Further specialization appears to occur within each of these broad domains. Each basal ganglia circuit originates in layer V of the cortex, and through a last step in the thalamus, returns to the cortex, not far from where the circuit started. Similarly arranged cortico-cerebello-thalamo-cortical loops also exist. Are such subcortical loops that are connected at both ends to the thalamocortical system necessarily part of the main thalamocortical complex? According to computer models, adding a loop to a large complex does not necessarily produce a larger complex with higher Φ. Instead, the elements of the main complex and those of the loop form a joint complex that can only integrate the limited amount of information exchanged within the loop. Thus subcortical loops could implement specialized subroutines that influence the states of the main thalamocortical complex, even if the neural interactions within the loop remain informationally insulated from those of the main complex. Informationally insulated cortico-subcortical loops may thus constitute substrates for unconscious processes that influence conscious experience.

5.6. Slow-wave sleep

Even if the anatomic connectivity is the same, a change in a few key parameters may radically alter the Φ value of a complex, with corresponding consequences on consciousness. When the brain enters the deepest stages of slow-wave sleep, consciousness is substantially reduced; yet neurons fire almost as intensely as in wakefulness. However, because of certain conductances modified by changes in neuromodulatory levels, neurons undergo periods of hyperpolarization, during which they are much further from their firing threshold. Computer simulations indicate that, if during such periods a part of the thalamocortical complex is perturbed in many different ways, the resulting states of the rest of the complex would be much less differentiated than during normal wakefulness. Although this conjecture has not yet been explicitly tested in vivo, if it were confirmed, it would indicate that the ability to integrate information is significantly reduced during slow-wave sleep.

5.7. Spatial and temporal scales

Values of Φ are dependent on both spatial and temporal scales that determine what counts as a state of the system. In general, the right "grain size" in both space and time should be the one at which Φ/t reaches a maximum. Biological considerations suggest that synchronous firing of heavily interconnected groups of neurons sharing inputs and outputs (e.g., cortical minicolumns) may produce significant effects in the rest of the brain, whereas asynchronous firing of various

combinations of individual neurons should not. Thus, Φ values may be higher when considering as elements cortical minicolumns rather than individual neurons, even if their number is lower. In regard to time, Φ values in the brain are likely to show a maximum between tens and hundreds of milliseconds. It is clear, for example, that if one were to perturb one half of the brain and examine what effects this produces on the other half, no perturbation would produce any effect whatsoever after just a tenth of a millisecond ($\Phi = 0$). However, after 100 milliseconds, for example, there is enough time for differential effects to be manifested. On the other hand, given the duration of conduction delays and of postsynaptic currents, longer intervals will not increase Φ values. Thus the time scale of neurophysiologic interactions needed to integrate information among distant cortical regions is consistent with that required by the microgenesis of conscious experience.

6. A few corollaries and predictions

Several corollaries and predictions derive from viewing and measuring consciousness as the amount of information that can be integrated within a complex of elements. Some of them are straightforward. For instance, stimulation and recording studies should indicate that groups of neurons in the thalamocortical system form a large complex of high Φ, whereas those in the cerebellum do not. Moreover, they should indicate that effective information among various cortical areas is high during waking and REM sleep and low in slow-wave sleep. Other predictions are less intuitive. For example, as long as neurons belong to a complex with high Φ, they should contribute to conscious experience whether they are firing or not, just as the sound of an orchestra is specified both by the instruments that are playing and by those that are silent. This is because the level of consciousness is a property of a system—a complex with high Φ—not of a state. Another prediction is that the same neurons can contribute or not to consciousness depending on whether they are part of the main thalamocortical complex. Yet another prediction states that when a task becomes automated, its execution fades from consciousness because the neural circuits that carry it become informationally insulated.

At a more general level, this approach makes it easy to see that consciousness provides an adaptive advantage and may have evolved precisely because it corresponds to the ability to rapidly integrate a lot of information. If such information is about the environment, the implication is that the more an animal is conscious, the larger the number of variables it can take into account jointly to guide its behavior. Also, in this view, consciousness would not be an all-or-none

property, but rather would increase in proportion to a system's ability to integrate information rapidly. Indeed, this approach provides a principled way of establishing, through the analysis of complexes, whether consciousness increases during development, whether and to what extent it is present in animals, and how much consciousness is left in neurologic conditions such as psychomotor seizures and akinetic mutism. Finally, an intriguing implication is that it should be possible to construct a conscious artifact, provided its physical basis is a complex having high Φ/t.

In summary, according to this theoretical framework, consciousness comes about when the parameters of a physical system are such that it can integrate a large amount of information over a short period of time. The level of consciousness can be determined, at least in principle, by evaluating $\Phi(S)/t$ for all subsets of a system at the spatial and temporal scale at which this quantity reaches a maximum. Such an evaluation identifies complexes—subsets of elements that can genuinely integrate information—and determines how much information they can integrate. At present, the validity of this theoretical framework rests on its ability to account, in a coherent manner, for some elementary but puzzling facts about consciousness and the brain. However, several experimental and theoretical developments are needed to establish whether it can provide a useful approach to the first problem of consciousness. Moreover, it remains to be seen whether the second problem of consciousness—what determines the way consciousness is—can be usefully addressed within the same framework. An interesting possibility is that, just as the amount of effective information that can be integrated within a complex would determine the level of consciousness, the space defined by the values of effective information among the elements of a complex would determine its content. Undoubtedly, a full understanding of how the brain generates consciousness remains a formidable task. However, if experimental investigations can usefully be complemented by principled theoretical approaches, it may not lay necessarily beyond the reach of science.

7. Further reading

Baars B J (1997): *Inside the theater of consciousness.* New York: Oxford University Press [MEDLINE]

Bachmann T (2000): *Microgenetic Approach to the Conscious Mind.* Amsterdam and Philadelphia: John Benjamins [MEDLINE]

Crick F, Koch C (2003): A framework for consciousness. *Nature Neuroscience* 2:119-126 [**MEDLINE**]

Dehaene S, Naccache L (2001): Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79:1-37 [**MEDLINE**]

Edelman GM, Tononi G (2000): *Consciousness: How Matter Becomes Imagination.* New York: Basic Books [MEDLINE]

Rees G, Kreiman G, Koch C (2001): Neural correlates of consciousness in humans. *Nature Reviews Neuroscience* 3:261-270 [**MEDLINE**]

Tononi G (2003): *Galileo and the photodiode*. Bari, Italy: Laterza [MEDLINE]

Zeman A (2003): *Consciousness: A User's Guide*. New Haven: Yale University Press [MEDLINE]

8. References

Shannon CE, Weaver W (1949): *The mathematical theory of communication.* Urbana: University of Illinois Press

Sporns O, Tononi G, et al. (2000): Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex* 10(2):127-141

Tononi G (2001): Information measures for conscious experience. *Archives Italiennes de Biologie* 139(4):367-371

Tononi G (2001): Information measures for conscious experience. *Archives Italiennes de Biologie* 139(4):367-371

Tononi G, Edelman GM (1998): Consciousness and complexity. *Science* 282 (5395):1846-1851

Tononi G, Sporns O, Edelman GM (1994): A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of the United States of America* 91(11):5033-5037

top