S. Laureys (Ed.)
Progress in Brain Research, Vol. 150
ISSN 0079-6123

CHAPTER 9

Consciousness, information integration, and the brain

Giulio Tononi*

Department of Psychiatry, University of Wisconsin, 6001 Research Park Blvd, Madison, WI 53719, USA

Abstract: Clinical observations have established that certain parts of the brain are essential for consciousness whereas other parts are not. For example, different areas of the cerebral cortex contribute different modalities and submodalities of consciousness, whereas the cerebellum does not, despite having even more neurons. It is also well established that consciousness depends on the way the brain functions. For example, consciousness is much reduced during slow wave sleep and generalized seizures, even though the levels of neural activity are comparable or higher than in wakefulness. To understand why this is so, empirical observations on the neural correlates of consciousness need to be complemented by a principled theoretical approach. Otherwise, it is unlikely that we could ever establish to what extent consciousness is present in neurological conditions such as akinetic mutism, psychomotor seizures, or sleepwalking, and to what extent it is present in newborn babies and animals. A principled approach is provided by the information integration theory of consciousness. This theory claims that consciousness corresponds to a system's capacity to integrate information, and proposes a way to measure such capacity. The information integration theory can account for several neurobiological observations concerning consciousness, including: (i) the association of consciousness with certain neural systems rather than with others; (ii) the fact that neural processes underlying consciousness can influence or be influenced by neural processes that remain unconscious; (iii) the reduction of consciousness during dreamless sleep and generalized seizures; and (iv) the time requirements on neural interactions that support consciousness.

Neuroscience and consciousness: facts and challenges

When addressing consciousness, two main problems need to be considered. The first problem is understanding the conditions that determine to what extent consciousness is present or absent. For example, why are changes of neural activity in thalamocortical regions so important for conscious experience, whereas changes in neural activity in cerebellar circuits are not, given that the number of neurons in the two structures is comparable? Or, why is consciousness strikingly reduced during deep slow-wave sleep, given that the average level of neuronal firing is similar to that of wakefulness?

The second problem is understanding the conditions that determine the specific way consciousness is. For example, what makes the activity of specific cortical areas contribute to specific dimensions of conscious experience — auditory cortex to sound, visual cortex to shapes, or 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 pain feels? Solving the first problem means that we would know to what extent a physical system can generate consciousness — the *quantity* or level of consciousness. Solving the second problem means that we would

*Corresponding author. Tel.: +1 (608) 263-6063; Fax: +1 (608) 2639340; E-mail: gtononi@wisc.edu

DOI: 10.1016/S0079-6123(05)50009-8

know what kind of consciousness it generates — the *quality* or content of consciousness.

The first problem is best considered by examining some well-established facts about the relationship between consciousness and the brain (the second problem is discussed in Tononi, 2004b). Each of these facts poses a serious challenge to our efforts to understand how consciousness comes about. Considered together, however, they strongly constrain the realm of possible answers.

1. Consciousness is produced by certain parts of the brain and not, or much less, by others. The parts that are essential are distributed within the thalamocortical system. Different areas of the thalamocortical system independently contribute different dimensions to conscious experience, and no single area is solely responsible for consciousness. What is special about this distributed network of thalamocortical circuits?
2. Other regions of the brain, such as the cerebellum, are not essential for consciousness, 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?
3. 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, while retinal cells can discriminate 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?
4. Neural activity in motor pathways is necessary to bring about the diverse behavioral responses that we usually associate with consciousness. However, such neural activity does not in itself contribute to consciousness. For example, patients with the locked-in syndrome, who are completely paralyzed except for the ability to gaze upward, are fully conscious. Similarly, we are paralyzed during dreams, but consciousness is not impaired by the absence of behavior. Even lesions of central motor areas do not impair consciousness. Why are we not conscious of what goes on in motor pathways?
5. 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. Moreover, some action sequences may be performed consciously when we first learn them, but fade from awareness when they become automatic. At the same time, their cortical substrates may shrink and shift to different circuits. Why are neural processes that take place automatically within cortico-subcortico-cortical circuits less conscious, and how do they become so?
6. Even within the thalamocortical system, many neural processes 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. Correspondingly, neurophysiological studies indicate that while the activity of certain cortical 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?

7. 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. Other neurological disconnection syndromes, as well as certain psychiatric dissociations, indicate that anatomical or functional disconnections among brain areas result in the shrinking or splitting of consciousness. What does this reveal about the neural substrate of consciousness?
8. On average, 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?
9. The firing of the same cortical neurons may correlate with consciousness at certain times, but not at other times. 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 irrespective of whether the animal reports seeing a figure or not. After 80–100 ms, 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 faced by biologists 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 have occurred. 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 review a theoretical approach according to which consciousness corresponds to the brain's ability to rapidly integrate information (Tononi, 2001, 2004a). The present review of the information integration theory of consciousness (IITC), which closely follows the original publications, comprises (i) an examination of phenomenology indicating that consciousness has to do with integrating information; (ii) a definition of what integrated information is and how it can be measured; (iii) an attempt at accounting for basic facts about consciousness and the brain; and (iv) some corollaries and predictions.

Phenomenology: consciousness as information integration

According to the IITC, perhaps the most important thing to realize about consciousness is that when one experiences a particular conscious state — say 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 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, and 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 that 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 a reduction of uncertainty among a number of alternatives (Shannon & Weaver, 1963). 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 dice 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. While 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 the IITC is that the information associated with the occurrence of a conscious state is not information from the perspective of an external observer, but *integrated information*. 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, i.e. 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 or how they sound in one's mind. Similarly, one cannot experience visual shapes independently of their color, or perceive the left half of the visual field of view independently of the right half. If one could, this would be tantamount to having two separate "centers" of consciousness. Separate centers of consciousness exist, of course, but then each is a different person with a different brain (or a different hemisphere of a split brain).

Finally, it is important to appreciate the characteristic spatio-temporal grain of consciousness. For example, psychophysical evidence indicates that sensory experiences require at least 100–200 ms to become progressively specified and stabilized. On the other hand, a single conscious moment cannot extend beyond 2–3 s. While it is arguable whether conscious experience unfolds more akin to a series of discrete snapshots or to a continuous flow, its time scale is certainly comprised between these lower and upper limits. Thus, a phenomenological analysis indicates that consciousness requires the integration of a large amount of information over a characteristic time scale.

Theory: measuring information integration

If consciousness corresponds to information integration, then a physical system should be able to generate consciousness to the extent that it can rapidly enter any 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, e.g. 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 dice, by using the entropy function, i.e. the weighted sum of the logarithm of the probability (p) of system states (s): $H = -\sum p(s)\log_2 p(s)$. 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 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 integrated. 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 it could easily be demonstrated by presenting it with all possible input signals. However, due to the absence of any physical interaction among the photodiodes, the

chip as such does not 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 information integration, it is essential to know whether a set of elements constitute a causally integrated system, or they 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 constituted of a set of elements (Tononi & Sporns, 2003; Tononi, 2004b). 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 hundreds of 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.

Consider now a subset S of elements taken from such a system, and the diagram of causal interactions among them (Fig. 1a). We will 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 within S, i.e. each state results from causal interactions within S. To do so, we divide S into two complementary parts A and B. We can now evaluate the responses of B that can be caused by all possible inputs originating from A. 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. In information-theoretical terms, we give maximum entropy to the outputs from A ($A^{H^{max}}$), i.e. we substitute its elements with independent noise sources, and we determine the entropy of the responses of B that can be caused by inputs from A. Specifically, we define the *effective information* between A and B as $EI(A \rightarrow B) = MI(A^{H^{max}}; B)$. Here $MI(A; B) = H(A) + H(B) - H(AB)$ stands for mutual information, a measure of the entropy or information shared between a source (A) and a target (B). Note that since A is substituted by independent noise sources, the entropy shared by B and A is necessarily due to causal effects of A on B. Moreover, $EI(A \rightarrow B)$ 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 $EI(B \rightarrow A)$ in general are not symmetric. For a given bipartition of a subset, then, the sum of the effective information for both directions is indicated as $EI(A \leftrightharpoons B) = EI(A \rightarrow B) + EI(B \rightarrow A)$. In summary, $EI(A \leftrightharpoons B)$ measures the repertoire of possible causal effects of A on B and of B on A.

On the basis of 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, we note that a subset S of elements cannot integrate any information if there is a way to partition S in two parts A and B such that $EI(A \leftrightharpoons B) = 0$ (Fig. 1b, vertical bipartition). In such a case we would obviously be dealing with at least two causally independent subsets, rather than with a single, integrated subset. This is exactly what would happen with the photodiodes making up the sensor of a digital camera: perturbing the state of some of the photodiodes would make no difference to the state of others. More generally, to measure the information integration capacity of a subset S, we should search for the bipartition(s) of S for which, after appropriate normalization, $EI(A \leftrightharpoons B)$ is lowest: its informational "weakest link", or *minimum information bipartition* ${}^{MIB}A \leftrightharpoons B$. The *information integration* for subset S, or $\Phi(S)$, is simply the (non-normalized) value of $EI(A \leftrightharpoons B)$ for its minimum information bipartition: $\Phi(S) = EI({}^{MIB}A \leftrightharpoons B)$. The symbol Φ is meant to indicate that the information (the vertical bar "I") is integrated within a single entity (the circle "O").

If $\Phi(S)$ is calculated for every possible subset S of a system, one can establish which subsets are actually capable of integrating information, and how much of it (Fig. 1c). After discarding all those subsets that are included in larger subsets having higher Φ (since they are merely parts of a larger whole), one is left with the *complexes* that make up the system. Specifically, a *complex* is a subset S

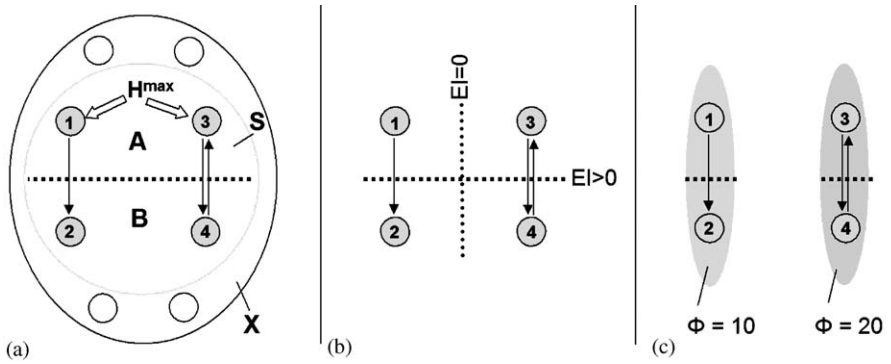

Fig. 1. Effective information, minimum information bipartition, and complexes. (a) Effective information. Shown is a single subset S of 4 elements ($\{1,2,3,4\}$, gray circle), forming part of a larger system X (black ellipse). This subset is bisected into A and B by a bipartition ($\{1,3\}/\{2,4\}$, indicated by the dotted gray line). Arrows indicate causally effective connections linking A to B and B to A across the bipartition (other connections may link both A and B to the rest of the system X). To measure EI(A → B), maximum entropy $H^{\max}$ is injected into the outgoing connections from A (corresponding to independent noise sources). The entropy of the states of B that is due to the input is then measured. Note that A can affect B directly through connections linking the two subsets, as well as indirectly via X. Applying maximum entropy to B allows one to measure EI(B → A). The effective information for this bipartition is EI(A ⇋ B) = EI(A → B) + EI(B → A). (b) Minimum information bipartition. For subset $S = \{1,2,3,4\}$, the horizontal bipartition $\{1,3\}/\{2,4\}$ yields a positive value of EI. However, the bipartition $\{1,2\}/\{3,4\}$ yields EI = 0 and is a minimum information bipartition (MIB) for this subset. The other bipartitions of subset $S = \{1,2,3,4\}$ are $\{1,4\}/\{2,3\}$, $\{1\}/\{2,3,4\}$, $\{2\}/\{1,3,4\}$, $\{3\}/\{1,2,4\}$, $\{4\}/\{1,2,3\}$, all with EI > 0. (c) Analysis of complexes. By considering all subsets of system X one can identify its complexes and rank them by the respective values of Φ — the value of EI for their minimum information bipartition. Assuming that other elements in X are disconnected, it is easy to see that Φ > 0 for subset $\{3,4\}$ and $\{1,2\}$, but Φ = 0 for subsets $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$, $\{1,2,3\}$, $\{1,2,4\}$, $\{1,3,4\}$, $\{2,3,4\}$, and $\{1,2,3,4\}$. Subsets $\{3,4\}$ and $\{1,2\}$ are not part of a larger subset having higher Φ, and therefore they constitute complexes. This is indicated schematically by having them encircled by a gray oval (darker gray indicates higher Φ). In order to identify complexes and their Φ(S) for systems with many different connection patterns, each system X was implemented as a stationary multidimensional Gaussian process such that values for effective information could be obtained analytically (for more details see Tononi and Sporns, 2003).

having Φ > 0 that is not included within a larger subset having higher Φ. For a complex, and only for a complex, it is appropriate to say that, when it enters a particular state out if its repertoire, it generates an amount of integrated information corresponding to its Φ value. Of the complexes that make up a given system, the one with the maximum value of Φ(S) is called the *main complex*. Some properties of complexes are worth pointing out. For example, 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. Also, the same element can belong to more than one complex, and complexes can overlap. One should also note that the Φ value of a complex is dependent on both spatial and temporal scales that determine what counts as a state of the underlying system. In general, the relevant spatial and temporal scales are those that jointly maximize Φ (Tononi, 2004b). In the case of the brain, the spatial elements and time scales that maximize Φ are likely to be local collections of neurons such as minicolumns and periods of time comprised between tens and hundreds of milliseconds, respectively.

In summary, a system can be analyzed to identify its complexes — those subsets of elements that can integrate information, and each complex will have an associated value of Φ, i.e. the amount of information it can integrate. To the extent that consciousness corresponds to the capacity to integrate information, complexes are the "subjects" of experience, being the locus where information can be integrated. Since information can only be integrated *within* a complex and not outside its boundaries, consciousness as information integration is necessarily subjective, private, and related to a single point of view or perspective (Tononi &

Edelman, 1998; Edelman & Tononi, 2000). It follows that elements that are part of a complex contribute to its conscious experience, while elements that are not part of it do not, even though they may be connected to it and exchange information with it through ports-in and ports-out.

Neuroscience: consciousness, information integration, and the brain

If consciousness corresponds to information integration, and if information integration 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?

Thalamocortical system. A well-functioning thalamocortical system is essential for consciousness (Plum, 1991), although opinions differ about the contribution of specific cortical areas (Tononi & Edelman, 1998; Zeman, 2001; Rees et al., 2002; Crick & Koch, 2003). Studies of comatose or vegetative patients indicate that a global loss of consciousness is usually caused by gray or white matter lesions that impair multiple sectors of the thalamocortical system (Adams et al., 2000; Laureys et al., 2002, 2004; Schiff et al., 2002). By contrast, selective lesions of individual thalamocortical areas impair different submodalities of conscious experience, such as the perception of color or of faces (Kolb & Whishaw, 1996). A global, persistent disruption of consciousness can also be produced by focal lesions of paramedian mesodiencephalic structures, which include the intralaminar thalamic nuclei (Schiff, 2004). Most likely, such focal lesions are catastrophic because the strategic location and connectivity of paramedian structures ensure that distributed cortico-thalamic loops can work together as a system. Electrophysiological and imaging studies also indicate that neural activity that correlates with conscious experience is widely distributed over the cortex (Tononi et al., 1998; Srinivasan et al., 1999; McIntosh et al., 1999, 2003; Rees et al., 2002). It would seem, therefore, that the neural substrate of consciousness is a distributed thalamocortical network, and that there is no single cortical area where it all comes together.

The fact that consciousness as we know it is generated by the thalamocortical system fits well with the IITC, since what we know about its organization appears ideally suited to the integration of information. On the information side, the thalamocortical system comprises a large number of elements that are functionally specialized, becoming activated in different circumstances (Zeki, 1993). Thus, the cerebral cortex is subdivided into systems dealing with different functions, such as vision, audition, motor control, planning, and many others. Each system in turn is subdivided into specialized areas, for example, different visual areas are activated by shape, color, and motion. Within an area, different groups of neurons are further specialized, e.g. by responding to different directions of motion. On the integration side, the specialized elements of the thalamocortical system are linked by an extended network of intra- and inter-areal connections that permit rapid and effective interactions within and between areas (Engel et al., 2001). In this way, thalamocortical neuronal groups are kept ready to respond, at multiple spatial and temporal scales, to activity changes in nearby and distant thalamocortical areas. As suggested by the regular finding of neurons showing multimodal responses that change depending on the context (Cohen & Andersen, 2002; Ekstrom et al., 2003), the capacity of the thalamocortical system to integrate information is probably greatly enhanced by non-linear switching mechanisms, such as gain modulation or synchronization, that can modify mappings between brain areas dynamically (Pouget et al., 2002; Tononi et al., 1992). In summary, the thalamocortical system is organized in a way that appears to emphasize at once both functional specialization and functional integration.

As shown by computer simulations, systems of neural elements whose connectivity jointly satisfies the requirements for functional specialization and

functional integration are well suited to integrating information. Fig. 2a shows a representative connection matrix obtained by optimizing for Φ starting from random connection weights. A graph-theoretical analysis indicates that connection matrices yielding the highest values of information integration (Φ = 74 bits) share two key characteristics. First, connection patterns are different for different elements, ensuring functional specialization. Second, all elements can be reached from all other elements of the network, ensuring functional integration. Thus, simulated systems having maximum Φ appear to require both functional specialization and functional integration. In fact, if

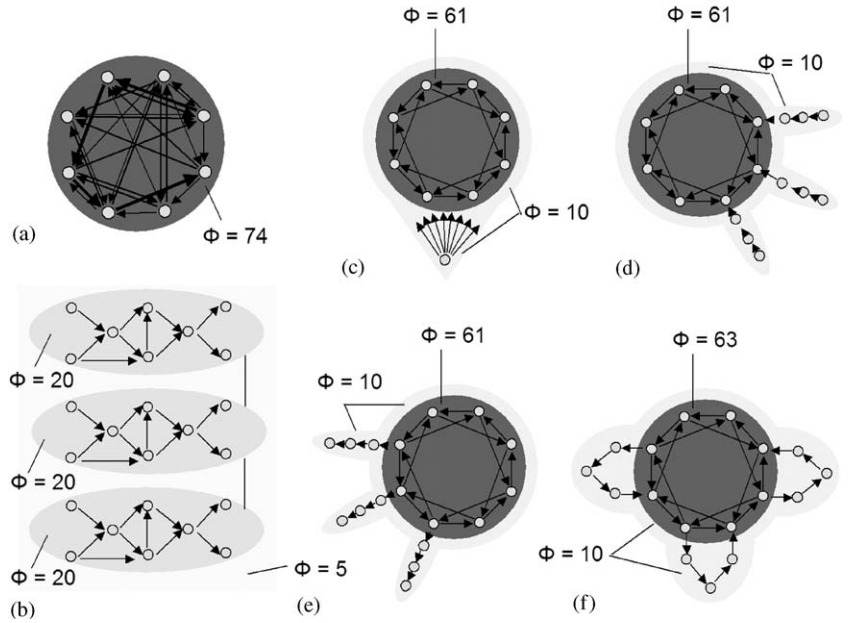

Fig. 2. Information integration for prototypical neural architectures. (a) Schematic of a cortical-like organization obtained by optimization of information integration. Shown is the causal interaction diagram for a network whose connection matrix was obtained by optimization for Φ (Φ = 74 bits). Note the heterogeneous arrangement of the incoming and outgoing connections: each element is connected to a different subset of elements, with different weights. Further analysis indicates that this network jointly maximizes functional specialization and functional integration among its eight elements, thereby resembling the anatomical organization of the thalamocortical system (Tononi and Sporns, 2003). (b) Schematic of a cerebellum-like organization. Shown are three modules of eight elements each, with many feed forward and lateral connections within each module but minimal connections among them. The analysis of complexes reveals three separate complexes with low values of Φ (Φ = 20 bits). There is also a large complex encompassing all the elements, but its Φ value is extremely low (Φ = 5 bits). (c) Schematic of the organization of a subcortical activating system. Shown is a single subcortical ''reticular'' element providing common input to the eight elements of a thalamocortical-like main complex (both specialized and integrated, Φ = 61 bits). Despite the diffuse projections from the reticular element on the main complex, the complex comprising all nine elements has a much lower value of Φ (Φ = 10 bits). (d) Schematic of the organization of afferent pathways. Shown are three short chains that stand for afferent pathways. Each chain connects to a port-in of a main complex having a high value of Φ (61 bits) that is thalamocortical-like (both specialized and integrated). Note that the afferent pathways and the elements of the main complex together constitute a large complex, but its Φ value is low (Φ = 10 bits). Thus, elements in afferent pathways can affect the main complex without belonging to it. (e) Schematic of the organization of efferent pathways. Shown are three short chains that stand for efferent pathways. Each chain receives a connection from a port-out of the thalamocortical-like main complex. Also in this case, the efferent pathways and the elements of the main complex together constitute a large complex, but its Φ value is low (Φ = 10 bits). (f) Schematic of the organization of cortico-subcortico-cortical loops. Shown are three short chains that stand for cortico-subcortico-cortical loops, which are connected to the main complex at both ports-in and ports-out. Again, the subcortical loops and the elements of the main complex together constitute a large complex, but its Φ value is low (Φ = 10 bits). Thus, elements in loops connected to the main complex can affect it without belonging to it. Note, however, that the addition of these three loops slightly increased the Φ value of the main complex (from Φ = 61 to Φ = 63 bits) by providing additional pathways for interactions among its elements.

functional specialization is lost by replacing the heterogeneous connectivity with a homogeneous one, or if functional integration is lost by rearranging the connections to form small modules, the value of Φ decreases considerably (Tononi & Sporns, 2003). Further simulations show that it is possible to construct a large complex of high Φ by joining smaller complexes through reciprocal connections. In the thalamocortical system, reciprocal connections linking topographically organized areas may be especially effective with respect to information integration. Thus, the coexistence of functional specialization and functional integration, epitomized by the thalamocortical system, is associated with high values of Φ.

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. According to the IITC, the reason lies with the organization of cerebellar connections, which is radically different from that of the thalamocortical system and is not well suited to information integration. 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 possible between distant patches (Bower, 2002; Cohen & Yarom, 1998). This suggests that cerebellar connections may not be organized so as to generate a large complex of high Φ, but rather to give rise to many small complexes each with a low value of Φ. Such an organization seems to be highly suited for both the learning and the rapid, effortless execution of informationally insulated subroutines.

This concept is illustrated in Fig. 2b, which shows a strongly *modular* network, consisting of three modules of eight strongly interconnected elements each. This network yields $\Phi = 20$ bits for each of its three modules, which form the system's three complexes. This example indicates that, irrespective of how many elements and connections are present in a neural structure, if that structure is organized in a strongly modular manner with little interactions among modules, complex size and Φ values are necessarily low. According to the IITC, this is the reason why these systems, although computationally very sophisticated, contribute little to consciousness. It is also the reason why there is no conscious experience associated with hypothalamic and brainstem circuits that regulate important physiological variables, such as blood pressure.

Activating systems. It has been known for a long time that lesions in the reticular formation of the brainstem can produce unconsciousness and coma. Conversely, stimulating the reticular formation can arouse a comatose animal and activate the thalamocortical system, making it ready to respond to stimuli (Moruzzi & Magoun, 1949). Groups of neurons within the reticular formation are characterized by diffuse projections to many areas of the brain. Many such groups release neuromodulators such as acetylcholine, histamine, noradrenaline, serotonin, dopamine, and glutamate (acting on metabotropic receptors) and can have extremely widespread effects on both neural excitability and plasticity (Steriade & McCarley, 1990). However, it would seem that the reticular formation, while necessary for the normal functioning of the thalamocortical system and therefore for the occurrence of conscious experience, may not contribute much in terms of specific dimensions of consciousness — it may work mostly like an external on-switch or as a transient booster of thalamocortical firing.

Such a role can be explained readily in terms of information integration. As shown in Fig. 2c, neural elements that have widespread and effective connections to a main complex of high Φ may nevertheless remain informationally excluded from it. Instead, they are part of a larger complex having a much lower value of Φ.

Cortical input systems. What we see usually depends on the activity patterns that occur in the retina and that are relayed to the brain. However, many observations suggest that retinal activity does not contribute directly to conscious experience. Retinal cells surely can discriminate light

from dark and convey that information to visual cortex, but their rapidly shifting firing patterns do not correspond well with what we perceive. For example, during blinks and eye movements retinal activity changes dramatically, but visual perception does not. The retina has a blind spot at the exit of the optic nerve where there are no photoreceptors, and it has low spatial resolution and no color sensitivity at the periphery of the visual field, but we are not aware of any of this. More importantly, lesioning the retina does not prevent conscious visual experiences. For example, a person who becomes retinally blind as an adult continues to have vivid visual images and dreams. Conversely, stimulating the retina during sleep by keeping the eyes open and presenting various visual inputs does not yield any visual experience and does not affect visual dreams. Why is it that retinal activity usually determines what we see through its action on thalamocortical circuits, but does not contribute directly to conscious experience?

As shown in Fig. 2d, adding or removing multiple, segregated incoming pathways to or from a main complex does not change the composition of the main complex, and causes little change in its Φ. While the incoming pathways do participate in a larger complex together with the elements of the main complex, the Φ value of this larger complex is very low, being limited by the effective information between each afferent pathway and its port-in at the main complex. Thus, input pathways providing powerful inputs to a complex add nothing to the information it integrates if their effects are entirely accounted for by ports-in.

Cortical output systems. Similar considerations apply to cortical output systems. In neurological practice, as well as in everyday life, we tend to associate consciousness with the presence of a diverse behavioral repertoire. For example, if we ask a lot of different questions and for each of them we obtain an appropriate answer, we generally infer that a person is conscious. Such a criterion is not unreasonable in terms of information integration, given that a wide behavioral repertoire is usually indicative of a large repertoire of internal states that is available to an integrated system. However, it appears that neural activity in motor pathways, which is necessary to bring about such diverse behavioral responses, does not in itself contribute to consciousness. For example, patients with the locked-in syndrome, who are completely paralyzed except for the ability to gaze upward, are fully conscious. Similarly, during dreams, consciousness is not impaired despite the absence of overt behavior. Even lesions of central motor areas do not impair consciousness.

Why is it that neurons in motor pathways, which can produce a large repertoire of different outputs and thereby relay a large amount of information about different conscious states, do not contribute directly to consciousness? As shown in Fig. 2e, adding or removing multiple, segregated outgoing pathways to or from a main complex does not change the composition of the main complex, and its Φ value. Like incoming pathways, outgoing pathways do participate in a larger complex together with the elements of the main complex, but the Φ value of this larger complex is very low, being limited by the effective information between each port-out of the main complex and its effector targets.

Basal ganglia and cortico-subcortical loops. Another set of neural structures that may not contribute directly to conscious experience are subcortical structures such as the basal ganglia. The basal ganglia contain many circuits arranged in parallel, some implicated in motor and oculomotor control, others, such as the dorsolateral prefrontal circuit, in cognitive functions, the lateral orbitofrontal and anterior cingulate circuits, in social behavior, motivation, and emotion (Alexander et al., 1990). 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 (Middleton & Strick, 2000). Similarly arranged cortico-cerebello-thalamo-cortical loops also exist. Why is it that such complicated neural structures, which are tightly connected to the thalamocortical system at both ends, do not seem to contribute directly to conscious experience?

As shown in Fig. 2f, the addition of many parallel cycles generally does not change the composition of the main complex, although Φ values can

be altered. Instead, the elements of the main complex and of the connected cycles form a joint complex that can only integrate the limited amount of information exchanged within each cycle. Thus, subcortical cycles or loops implement specialized subroutines that are capable of influencing the states of the main thalamocortical complex without joining it. Such informationally insulated cortico-subcortical loops could constitute the neural substrates for many unconscious processes that can affect and be affected by conscious experience (Baars, 1988; Tononi, 2004a, b). It is likely that new informationally insulated loops can be created through learning and repetition. For example, when first performing a new task, we are conscious of every detail of it, we make mistakes, are slow, and must make an effort. When we have learned the task well, we perform it better, faster, and with less effort, but we are also less aware of it. As suggested by imaging results, a large number of neocortical regions are involved when we first perform a task. With practice, activation is reduced or shifts to different circuits (Raichle, 1998). According to the IITC, during the early trials, performing the task involves many regions of the main complex, while later certain aspects of the task are delegated to neural circuits, including subcortical ones, that are informationally insulated.

Cortical loops. Even within the thalamocortical system proper, a substantial proportion of neural activity does not appear to contribute directly to conscious experience. For example, what we see and hear requires elaborate computational processes dealing with figure-ground segregation, depth perception, object recognition, and language parsing, many of which take place in the thalamocortical system. Yet we are not aware of all this diligent buzzing: we just *see* objects, separated from the background and laid out in space, and know what they are, or *hear* words, nicely separated from each other, and know what they mean. As an example, take binocular rivalry, where the two eyes view two different images, but we perceive consciously just one image at a time, alternating in sequence. Recordings in monkeys have shown that the activity of visual neurons in certain cortical areas, such as the inferotemporal cortex, follows faithfully what the subject perceives consciously. However, in other areas, such as primary visual cortex, there are many neurons that respond to the stimulus presented to the eye, whether or not the subject is perceiving it (Logothetis et al., 1996). Neuromagnetic studies in humans have shown that neural activity correlated with a stimulus that is not being consciously perceived can be recorded in many cortical areas, including the front of the brain (Srinivasan et al., 1999). Why does the firing of many cortical neurons, carrying out the computational processes that enable object recognition (or language parsing), not correspond to anything conscious?

The situation is similar on the executive side of consciousness. When we plan to do or say something, we are vaguely conscious of what we intend, and presumably these intentions are reflected in specific firing patterns of certain neuronal groups. Our vague intentions are then translated almost miraculously into the right words, strung together to form a syntactically correct sentence that conveys what we meant to say. And yet again, we are not at all conscious of the complicated processing that is needed to carry out our intentions, much of which takes place in the cortex. What determines whether the firing of neurons within the thalamocortical system, contributes directly to consciousness or not? According to the IITC, the same considerations that apply to input and output circuits and to cortico-subcortico-cortical loops also apply to circuits and loops contained entirely within the thalamocortical system. Thus, the theory predicts that activity within certain cortical circuits does not contribute to consciousness because such circuits implement informationally insulated loops that remain outside the main thalamocortical complex. At this stage, however, it is hard to say precisely which cortical circuits may be informationally insulated. Are primary sensory cortices organized like massive afferent pathways to a main complex "higher up" in the cortical hierarchy? Is much of prefrontal cortex organized like a massive efferent pathway? Do certain cortical areas, such as those belonging to the dorsal visual stream, remain partly segregated from the main complex? Do interactions *within* a cortico-thalamic minicolumn qualify as intrinsic

mini-loops that support the main complex without being part of it? Unfortunately, answering these questions and properly testing the predictions of the theory requires a much better understanding of cortical neuroanatomy than is presently available (Ascoli, 1999).

Anatomical and functional disconnections. 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 non-dominant hemisphere also appears to be conscious, although it loses some important abilities. Some information, e.g. emotional arousal, seems to be shared across the hemispheres, probably owing to subcortical common inputs. As illustrated by simple computer models (Tononi, 2004b), a "callosal" cut produces, out of 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 to 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.

In addition to anatomical disconnections, functional disconnections may also lead to a restriction of the neural substrate of consciousness. Functional disconnections between certain parts of the brain and others may play a role in neurological neglect phenomena, may underlie psychiatric conversion and dissociative disorders, may occur during dreaming, and may be implicated in conditions such as hypnosis. It is also possible that certain attentional phenomena may correspond to changes in the neural substrate of consciousness. For example, when one is absorbed in thought, or focused exclusively on a given sensory modality, such as vision, the neural substrate of consciousness may not be the same as when we are diffusely monitoring the environment. Phenomena such as the attentional blink, where a fixed sensory input may at times make it to consciousness and at times not, may also be due to changes in functional connectivity: access to the main thalamocortical complex may be enabled or not based on dynamics intrinsic to the complex (Dehaene et al., 2003). Phenomena such as binocular rivalry may also be related, at least in part, to dynamic changes in the composition of the main thalamocortical complex caused by transient changes in functional connectivity (Lumer, 1998). Computer simulations confirm that functional disconnection can reduce the size of a complex and reduce its capacity to integrate information (Tononi, 2004b). While it is not easy to determine, at present, whether a particular group of neurons is excluded from the main complex because of hard-wired anatomical constraints, or is transiently disconnected due to functional changes, the set of elements underlying consciousness is not static, but can be considered to form a "*dynamic complex*" or "*dynamic core*" (Tononi & Edelman, 1998).

Slow-wave sleep. If neuroanatomical organization is the key in enabling information integration and thereby consciousness, neurophysiological parameters are no less important. A case in point is provided by sleep, perhaps the most familiar and yet striking alteration of consciousness. Upon awakening from dreamless sleep, we have the peculiar impression that for a while we were not there at all nor, as far as we are concerned, was the rest of the world. This everyday observation tells us vividly that consciousness can be gained and lost, grow and shrink. Indeed, if we did not sleep, it might be hard to imagine that consciousness is not a given, but depends somehow on the way our brain is functioning. The loss of consciousness between falling asleep and waking up is relative, rather than absolute (Hobson et al., 2000). Thus, careful studies of mental activity reported immediately after awakening have shown that some degree of consciousness is maintained during much of sleep. Many awakenings, especially from rapid eye movement (REM) sleep, yield dream reports, and dreams can be at times as vivid and intensely conscious as waking experiences. Dream-like consciousness also occurs during various phases of slow-wave sleep, especially at sleep onset and during the last part of the night. Nevertheless, a

certain proportion of awakenings do not yield any dream report, suggesting a marked reduction of consciousness. Such ''empty'' awakenings typically occur during the deepest stages of slow wave sleep (stages 3 and 4), especially during the first half of the night.

Which neurophysiological parameters are responsible for the remarkable changes in the quantity and quality of conscious experience that occur during sleep? We know for certain that the brain does not simply shut off during sleep. During REM sleep, for example, neural activity is as high, if not higher, than during wakefulness, and EEG recordings show low-voltage fast-activity. This EEG pattern is known as ''activated'' because cortical neurons, being steadily depolarized and close to their firing threshold, are ready to respond to incoming inputs. Given these similarities, it is perhaps not surprising that consciousness should be present during both states. Changes in the quality of consciousness, however, do occur, and they correspond closely to relative changes in the activation of different brain areas (Hobson et al., 2000).

During slow wave sleep, average firing rates of cortical neurons are also similar to those observed during quiet wakefulness. However, due to changes in the level of certain neuromodulators, virtually all cortical neurons engage in slow oscillations at around 1 Hz, which are reflected in slow waves in the EEG (Steriade, 1997). Slow oscillations consist of a depolarized phase, during which the membrane potential of cortical neurons is close to firing threshold and spontaneous firing rates are similar to quiet wakefulness, and of a hyperpolarized phase, during which neurons become silent and are further away from firing threshold. From the perspective of information integration, a reduction in the readiness to respond to stimuli during the hyperpolarization phase of the slow oscillation would imply a reduction of consciousness. It would be as if we were watching very short fragments of a movie interspersed with repeated unconscious ''blanks'' in which we cannot see, think, or remember anything, and therefore have little to report. A similar kind of unreadiness to respond, associated with profound hyperpolarization, is found in deep anesthesia, another condition were consciousness is impaired.

From the perspective of information integration, a reduction of consciousness during certain phases of sleep would occur even if the brain remained capable of responding to perturbations, provided its response were to lack differentiation. This prediction is borne out by detailed computer models of a portion of the visual thalamocortical system (Hill & Tononi, in preparation). According to these simulations, in the waking mode different perturbations of the thalamocortical network yield specific responses. In the sleep mode, instead, the network becomes bistable. Specific effects of different perturbations are quickly washed out and their propagation impeded: the whole network transitions into the depolarized or into the hyperpolarized phase of the slow oscillation — a stereotypic response that is observed irrespective of the particular perturbation. And of course, this bistability is also evident in the spontaneous behavior of the network: during each slow oscillation, cortical neurons are either all firing or all silent, with little freedom in between. In summary, these simulations indicate that, even if the anatomical connectivity of a complex stays the same, a change in key parameters governing the readiness of neurons to respond, and the differentiation of their responses may alter radically the Φ value of the complex, with corresponding consequences on consciousness. Further simulations indicate that the capacity to integrate information is also reduced if neural activity is extremely high and near-synchronous, due to a dramatic decrease in the available degrees of freedom (Tononi, unpublished results). This reduction in degrees of freedom could be the reason why consciousness is reduced or eliminated in absence seizure and other conditions characterized by hypersynchronous neural activity.

Conscious experience and time. Consciousness not only requires a neural substrate with appropriate anatomical structure and appropriate physiological parameters: it also needs time. For example, studies of how a percept is progressively specified and stabilized indicate that it takes up to 100–200 ms to develop a fully formed sensory experience, and that the surfacing of a conscious thought may take even longer (Bachmann, 2000). Experiments in which the somatosensory areas of the cerebral cortex were

stimulated directly indicate that low intensity stimuli must be sustained for up to 500 ms to produce a conscious sensation (Libet, 1982). Multiunit recordings in the primary visual cortex of monkeys show that, after a stimulus is presented, the firing rate of many neurons increases irrespective of whether the animal reports seeing a figure or not. After 80–100 ms, however, their discharge accurately predicts the conscious detection of the figure. Thus, the firing of the same cortical neurons may correlate with consciousness at certain times, but not at other times (Lamme & Roelfsema, 2000). What determines when the firing of the same cortical neurons contributes to conscious experience and when it does not? And why does it take up to hundreds of milliseconds before a conscious experience is generated?

The IITC predicts that the time requirement for the generation of conscious experience in the brain emerge directly from the time requirements for the buildup of effective interactions among the elements of the thalamocortical main complex. As mentioned above, if one were to perturb half of the elements of the main complex for less than a millisecond, no perturbations would produce any effect on the other half within this time window, and Φ would be equal to zero. After, say, 100 ms, however, there is enough time for differential effects to be manifested, and Φ should grow. Thus, the time scale of neurophysiological interactions needed to integrate information among distant cortical regions appears to be consistent with that required by psychophysical observations (microgenesis), by stimulation experiments, and by recording experiments.

Comparisons and conclusions

The examples discussed above show that the IITC can account, in a coherent manner, for several puzzling facts about consciousness and the brain. How does the theory compare with other approaches to the neurobiology of consciousness, and what are some of its implications and predictions?

Few neuroscientists have devoted an organized body of work to the neural substrates of consciousness. Edelman (1989) was among the first to propose that consciousness should be addressed fully within a neurobiological framework. In several publications (Edelman, 1989, 2003), Edelman has maintained that consciousness requires reentrant interactions between posterior networks involved in perceptual categorization, and anterior-limbic networks involved in ''value-category'' memory, which result in a kind of ''remembered present''. This view represents an extension to consciousness of a more general, selectionist approach to brain function (Edelman, 1987). Key ideas are that it is useful to distinguish between primary and higher-order consciousness, that the substrate of consciousness is highly distributed and variable, and that consciousness requires a body and the interaction of the organism with an environment.

Crick and Koch were also among the first to advocate a research program aimed at identifying in progressively greater detail the neural correlates of consciousness (Crick & Koch, 1990). Their proposals are guided primarily by empirical considerations. Over the years, they have made several suggestions, ranging from the role of 40 Hz oscillations in binding different conscious attributes, to suggesting that only a small subset of neurons is associated with consciousness, to the idea that neurons associated with consciousness must project directly to prefrontal cortex, and that neurons in primary visual cortex do not contribute to consciousness (Crick & Koch, 1995, 1998). More recently, they have to some extent enlarged their scope and suggested that the substrate of consciousness may be ''coalitions'' of neurons, both in the front and the back of the cortex, which compete to establish some metastable, strong firing pattern that explicitly represents information and can guide action (Crick & Koch, 2003). Related ideas are that higher cortical areas as well as attention can strongly modulate the strength of conscious coalitions, that there is a penumbra of neural activity that gives ''meaning'' to conscious firing patterns, and that there are ''zombie'' neural systems that are fast but unconscious.

Dehaene and Changeux have taken as their starting point the global workspace theory (Dehaene and Naccache, 2001), elaborated most extensively in a cognitive context by Baars (1988). They have singled out, as experimentally more tractable, the

notion of global access — the idea that a "piece of information" encoded in the firing of a group of neurons becomes conscious if it is "broadcast" widely, so that a large part of the brain has access to it. That is, the same information can be conscious or not depending on the size of the audience. This formulation translates, in plausible neural terms, the key insight of global workspace theory, exemplified by the theater (or TV) metaphor: a message becomes conscious when it becomes accessible to a large audience (it goes on stage), but not if it remains private. Key ideas are that global workspace neurons, characterized by their ability to send and receive projections from many distant areas through long-range excitatory fibers, are especially concentrated in prefrontal, anterior cingulate, and parietal areas, that neurons must be actively firing (broadcasting) to contribute to consciousness, that access to consciousness is an all-or-none phenomenon, requiring the nonlinear "ignition" of global workspace neurons, and that higher areas play a role in "mobilizing" lower areas into the global workspace.

There are both similarities and differences between the IITC and neurobiological frameworks such as those just described. Not surprisingly, there is broad convergence on certain key facts: that consciousness is generated by distributed thalamocortical networks; reentrant interactions among multiple cortical regions are important; that the mechanisms of consciousness and attention overlap but are not the same, and that there are many "unconscious" neural systems. Of course, different approaches may emphasize different aspects. However, at the present stage these differences are not crucial, and fluctuate with the pendulum of experimental evidence.

The main differences lie elsewhere. Unlike other approaches, the IITC addresses the so-called hard problem (Chalmers, 1996) head-on. It takes its start from phenomenology and, by making a critical use of thought experiments, argues that subjective experience is integrated information. Therefore, any physical system will have subjective experience to the extent that it is capable of integrating information. In this view, experience, i.e. information integration, is a fundamental quantity, just as mass or energy. Other approaches avoid the hard problem and do not take a theoretical stand concerning the fundamental nature of experience, restricting themselves to the empirical investigation of its neural correlates.

The IITC takes a precise view about information integration, offering a general theoretical definition and a way to measure it as the Φ value of a complex. In other approaches, including the ones inspired by the global workspace metaphor, the notion of information is not well defined. For example, it is often assumed loosely that the firing of specific thalamocortical elements (e.g. those for red) conveys some specific information (e.g. that there is something red), and that such information becomes conscious if it is disseminated widely. However, just like a retinal cell or a photodiode, a given thalamocortical element has no information about whether what made it fire was a particular color rather than a shape, a visual stimulus rather than a sound, or a sensory stimulus rather than a thought. All it knows is whether it fired or not, just as each receiving element only knows whether it received an input or not. Thus, the information specifying "red" cannot possibly be in the message conveyed by the firing of any neural element, whether it is broadcasting widely or not. According to the IITC, that information resides instead in the reduction of uncertainty occurring when a whole complex enters one out of a large number of available states. Moreover, within a complex, both active and inactive neurons count, just as the sound of an orchestra is specified both by the instruments that are playing and by those that are silent. In short, what counts is how much information is generated, and not how widely it is disseminated.

By arguing that subjective experience corresponds to a system's capacity to integrate information, and by providing a mathematical definition of information integration, the IITC can go on to show that several observations concerning the neural substrate of consciousness fall naturally into place. Other approaches generally propose a provisional list of neural ingredients that appear to be important, such as synchronization or widespread broadcasting, without providing a principled explanation of why they would be important or whether they would be always necessary. For example, synchronization is usually an indication that the elements of the complex are capable of interacting efficiently, but

is neither necessary nor sufficient for consciousness: there can be strong synchronization with little consciousness (absence seizures) as well as consciousness with little synchronization (as indicate by unit recordings in higher-order visual areas). Or there can be extremely widespread "broadcasting", as exemplified most dramatically by the diffuse projections of neuromodulatory systems, yet lesion, stimulation and recording experiments do not suggest any specific contribution to specific dimensions of consciousness.

The IITC also predicts that consciousness depends exclusively on the ability of a system to integrate information, whether or not it has a strong sense of self, language, emotion, a body, or is immersed in an environment, contrary to some common intuitions. This prediction is consistent with the preservation of consciousness during REM sleep, when both input and output signals from and to the body, respectively are markedly reduced. Transient inactivation of brain areas mediating the sense of self, language, and emotion could assess this prediction in a more cogent manner. Nevertheless, the theory recognizes that these same factors are important historically because they favor the development of neural circuits forming a main complex of high Φ. For example, the ability of a system to integrate information grows as that system incorporates statistical regularities from its environment and learns (Tononi et al., 1996). In this sense, the emergence of consciousness in biological systems is predicated on a long evolutionary history, on individual development, and on experience-dependent changes in neural connectivity.

Finally, the IITC says that the presence and extent of consciousness can be determined, in principle, also in cases in which we have no verbal report, such as infants or animals, or in neurological conditions such as akinetic mutism, psychomotor seizures, and sleepwalking. In practice, of course, measuring Φ accurately in such systems will not be easy, but approximations and informed guesses are certainly conceivable. The IITC also implies that consciousness is not an all-or-none property, but increases in proportion to a system's ability to integrate information. In fact, any physical system capable of integrating information would have some degree of experience, irrespective of the stuff of which it is made.

At present, the validity of this theoretical framework and the plausibility of its implications rest on its ability to account, in a coherent manner, for some basic phenomenological observations and for some elementary but puzzling facts about consciousness and the brain. Experimental developments, especially of ways to concurrently stimulate and record the activity of broad regions of the brain, should permit stringent tests of some of the theory's predictions. Equally important will be the development of realistic, large-scale models of the anatomical organization of the brain. These models should allow a more rigorous measurement of how the capacity to integrate information relates to different brain structures and certain neurophysiological parameters (Tononi et al., 1992; Lumer et al., 1997; Hill & Tononi, 2005).

References

Adams, J.H., Graham, D.I. and Jennett, B. (2000) The neuropathology of the vegetative state after an acute brain insult. Brain, 123(Pt 7): 1327–1338.

Alexander, G. E., Crutcher, M. D. and DeLong, M. R. (1990) Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. Progress in Brain Research, Vol. 85, Elsevier, Amsterdam, pp. 119–146.

Ascoli, G.A. (1999) Progress and perspectives in computational neuroanatomy. Anat. Rec., 257: 195–207.

Baars, B.J. (1988) A Cognitive Theory of Consciousness. Cambridge University Press, New York.

Bachmann, T. (2000) Microgenetic Approach to the Conscious Mind. John Benjamins Pub. Co, Amsterdam, Philadelphia.

Bower, J.M. (2002) The organization of cerebellar cortical circuitry revisited: Implications for function. Ann. N Y Acad. Sci., 978: 135–155.

Chalmers, D.J. (1996) The Conscious Mind: In Search of a Fundamental Theory, Philosophy of Mind Series. Oxford University Press, New York.

Cohen, D. and Yarom, Y. (1998) Patches of synchronized activity in the cerebellar cortex evoked by mossy-fiber stimulation: Questioning the role of parallel fibers. Proc. Natl. Acad. Sci. USA, 95: 15032–15036.

Cohen, Y.E. and Andersen, R.A. (2002) A common reference frame for movement plans in the posterior parietal cortex. Nat. Rev. Neurosci., 3: 553–562.

Crick, F. and Koch, C. (1990) Some reflections on visual awareness. Cold Spring Harbor Symposia on Quantitative Biology, 55: 953–962.

Crick, F. and Koch, C. (1995) Are we aware of neural activity in primary visual cortex? Nature, 375: 121–123.
Crick, F. and Koch, C. (1998) Consciousness and neuroscience. Cereb. Cortex, 8: 97–107.
Crick, F. and Koch, C. (2003) A framework for consciousness. Nat. Neurosci., 6: 119–126.
Dehaene, S. and Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. Cognition, 79: 1–37.
Dehaene, S., Sergent, C. and Changeux, J.P. (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proc. Natl. Acad. Sci. USA, 100: 8520–8525.
Edelman, G.M. (1987) Neural Darwinism: The Theory of Neuronal Group Selection. BasicBooks, Inc, New York.
Edelman, G.M. (1989) The Remembered Present: A Biological Theory of Consciousness. BasicBooks, Inc, New York.
Edelman, G.M. (2003) Naturalizing consciousness: A theoretical framework. Proc. Natl. Acad. Sci. USA, 100: 5520–5524.
Edelman, G.M. and Tononi, G. (2000) A Universe of Consciousness: How Matter Becomes Imagination. Basic Books, New York.
Ekstrom, A.D., Kahana, M.J., Caplan, J.B., Fields, T.A., Isham, E.A., Newman, E.L. and Fried, I. (2003) Cellular networks underlying human spatial navigation. Nature, 425: 184–188.
Engel, A.K., Fries, P. and Singer, W. (2001) Dynamic predictions: Oscillations and synchrony in top-down processing. Nat. Rev. Neurosci., 2: 704–716.
Hill, S. and Tononi, G. (2005) Modeling sleep and wakefulness in the thalamocortical system. J. Neurophysiol., 93: 1671–1698.
Hobson, J.A., Pace-Schott, E.F. and Stickgold, R. (2000) Dreaming and the brain: Toward a cognitive neuroscience of conscious states. Behav. Brain Sci., 23: 793–842 discussion 904–1121.
Kolb, B. and Whishaw, I.Q. (1996) Fundamentals of Human Neuropsychology. WH Freeman, New York.
Lamme, V.A. and Roelfsema, P.R. (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci., 23: 571–579.
Laureys, S., Antoine, S., Boly, M., Elincx, S., Faymonville, M.E., Berre, J., Sadzot, B., Ferring, M., De Tiege, X., van Bogaert, P., Hansen, I., Damas, P., Mavroudakis, N., Lambermont, B., Del Fiore, G., Aerts, J., Degueldre, C., Phillips, C., Franck, G., Vincent, J.L., Lamy, M., Luxen, A., Moonen, G., Goldman, S. and Maquet, P. (2002) Brain function in the vegetative state. Acta. Neurol. Belg., 102: 177–185.
Laureys, S., Owen, A.M. and Schiff, N.D. (2004) Brain function in coma, vegetative state, and related disorders. Lancet Neurol., 3: 537–546.
Libet, B. (1982) Brain stimulation in the study of neuronal functions for conscious sensory experiences. Human Neurobiol., 1: 235–242.
Logothetis, N.K., Leopold, D.A. and Sheinberg, D.L. (1996) What is rivalling during binocular rivalry? Nature, 380: 621–624.
Lumer, E.D. (1998) A neural model of binocular integration and rivalry based on the coordination of action-potential timing in primary visual cortex. Cereb. Cortex, 8: 553–561.
Lumer, E.D., Edelman, G.M. and Tononi, G. (1997) Neural dynamics in a model of the thalamocortical system.1. Layers, loops and the emergence of fast synchronous rhythms. Cereb. Cortex, 7: 207–227.
McIntosh, A.R., Rajah, M.N. and Lobaugh, N.J. (1999) Interactions of prefrontal cortex in relation to awareness in sensory learning. Science, 284: 1531–1533.
McIntosh, A.R., Rajah, M.N. and Lobaugh, N.J. (2003) Functional connectivity of the medial temporal lobe relates to learning and awareness. J. Neurosci., 23: 6520–6528.
Middleton, F.A. and Strick, P.L. (2000) Basal ganglia and cerebellar loops: Motor and cognitive circuits. Brain Res. Brain Res. Rev., 31: 236–250.
Moruzzi, G. and Magoun, H.W. (1949) Brain stem reticular formation and activation of the EEG. Electroencephalog. Clin. Neurophysiol., 1: 455–473.
Plum, F. (1991) Coma and related global disturbances of the human conscious state. In: Peters A. and Jones E.G. (Eds.) Normal and Altered States of Function, Vol. 9. Plenum Press, New York, pp. 359–425.
Pouget, A., Deneve, S. and Duhamel, J.R. (2002) A computational perspective on the neural basis of multisensory spatial representations. Nat. Rev. Neurosci., 3: 741–747.
Raichle, M.E. (1998) The neural correlates of consciousness: An analysis of cognitive skill learning. Philos. Trans. R. Soc. Lond. B. Biol. Sci., 353: 1889–1901.
Rees, G., Kreiman, G. and Koch, C. (2002) Neural correlates of consciousness in humans. Nat. Rev. Neurosci., 3: 261–270.
Schiff, N.D. (2004). The neurology of impaired consciousness: Challenges for cognitive neuroscience. In: Gazzaniga, M. (Ed.), The Cognitive Neurosciences. 3rd ed. Cambridge, MA: MIT Press.
Schiff, N.D., Ribary, U., Moreno, D.R., Beattie, B., Kronberg, E., Blasberg, R., Giacino, J., McCagg, C., Fins, J.J., Llinas, R. and Plum, F. (2002) Residual cerebral activity and behavioural fragments can remain in the persistently vegetative brain. Brain, 125: 1210–1234.
Shannon, C.E. and Weaver, W. (1963) The Mathematical Theory of Communication. University of Illinois Press, Urbana.
Srinivasan, R., Russell, D.P., Edelman, G.M. and Tononi, G. (1999) Increased synchronization of neuromagnetic responses during conscious perception. J. Neurosci., 19: 5435–5448.
Steriade, M. (1997) Synchronized activities of coupled oscillators in the cerebral cortex and thalamus at different levels of vigilance. Cereb. Cortex, 7: 583–604.
Steriade, M. and McCarley, R.W. (1990) Brainstem Control of Wakefulness and Sleep. Plenum Press, New York.
Tononi, G. (2001) Information measures for conscious experience. Arch. Ital. Biol., 139: 367–371.
Tononi, G. (2004a) Consciousness and the brain: Theoretical aspects. In: Adelman G. and Smith B. (Eds.), Encyclopedia of Neuroscience. Elsevier, Amsterdam.

Tononi, G. (2004b) An information integration theory of consciousness. BMC Neurosci., 5: 42.
Tononi, G. and Edelman, G.M. (1998) Consciousness and complexity. Science, 282: 1846–1851.
Tononi, G. and Sporns, O. (2003) Measuring information integration. BMC Neurosci., 4: 31.
Tononi, G., Sporns, O. and Edelman, G.M. (1992) Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. Cereb. Cortex, 2: 310–335.
Tononi, G., Sporns, O. and Edelman, G.M. (1996) A complexity measure for selective matching of signals by the brain. Proc. Natl. Acad. Sci. USA, 93: 3422–3427.
Tononi, G., Srinivasan, R., Russell, D.P. and Edelman, G.M. (1998) Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. Proc. Natl. Acad. Sci. USA, 95: 3198–3203.
Zeki, S. (1993) A Vision of the Brain. Blackwell Scientific Publications, Oxford, Boston.
Zeman, A. (2001) Consciousness. Brain, 124: 1263–1289.