



# Consciousness and Complexity

Giulio Tononi\* and Gerald M. Edelman

Conventional approaches to understanding consciousness are generally concerned with the contribution of specific brain areas or groups of neurons. By contrast, it is considered here what kinds of neural processes can account for key properties of conscious experience. Applying measures of neural integration and complexity, together with an analysis of extensive neurological data, leads to a testable proposal—the dynamic core hypothesis—about the properties of the neural substrate of consciousness.

What is the neural substrate of conscious experience? While William James concluded that it was the entire brain (1), recent approaches have attempted to narrow the focus: are there neurons endowed with a special location or intrinsic property that are necessary and sufficient for conscious experience? Does primary visual cortex contribute to conscious experience? Are brain areas that project directly to prefrontal cortex more relevant than those that do not (2)? Although heuristically useful, these approaches leave a fundamental problem unresolved: How could the possession of some particular anatomical location or biochemical feature render some neurons so privileged that their activity gives rise to subjective experience? Conferring this property on neurons seems to constitute a category error, in the sense of ascribing to things properties they cannot have (3).

Here, we pursue a different approach. Instead of arguing whether a particular brain area or group of neurons contributes to consciousness or not, our strategy is to characterize the kinds of neural processes that might account for key properties of conscious experience. We emphasize two properties: conscious experience is integrated (each conscious scene is unified) and at the same time it is highly differentiated (within a short time, one can experience any of a huge number of different conscious states). We first consider neurobiological data indicating that neural processes associated with conscious experience are highly integrated and highly differentiated. We then provide tools for measuring integration (called functional clustering) and differentiation (called neural complexity) that are applicable to actual neural processes. This leads us to formulate operational criteria for

determining whether the activity of a group of neurons contributes to conscious experience. These criteria are incorporated into the dynamic core hypothesis, a testable proposal concerning the neural substrate of conscious experience (4).

## General Properties of Conscious Experience

Consciousness, as William James pointed out, is not a thing, but a process or stream that is changing on a time scale of fractions of seconds (1). As he emphasized, a fundamental aspect of the stream of consciousness is that it is highly unified or integrated.

*Integration.* Integration is a property shared by every conscious experience irrespective of its specific content: Each conscious state comprises a single “scene” that cannot be decomposed into independent components (5). Integration is best appreciated by considering the impossibility of conceiving of a conscious scene that is not integrated, that is, one which is not experienced from a single point of view. A striking demonstration is given by split-brain patients performing a spatial memory task in which two independent sequences of visuospatial positions were presented, one to the left and one to the right hemisphere (6). In these patients, each hemisphere perceived a separate, simple visual problem and the subjects were able to solve the double task well. Normal subjects could not treat the two independent visual sequences as independent, parallel tasks. Instead, they combined the visual information into a single conscious scene and into a single, large problem that was much more difficult to solve.

The unity of conscious experience is also evidenced by our inability to perform multiple tasks, unless some tasks are highly automatic and impinge less on consciousness. Moreover, we cannot make more than a single conscious decision within an interval of a few hundreds of milliseconds, the so-called psychological refractory period (7). Further-

more, we cannot be aware of two incongruent scenes at the same time, as indicated by the bistability of ambiguous figures and the phenomenon of perceptual rivalry (8). Unity also entails that conscious experience is private, that is, it is always experienced from a particular point of view and cannot fully be shared (1).

*Differentiation.* While each conscious state is an integrated whole, perhaps the most remarkable property of conscious experience is its extraordinary differentiation or complexity. The number of different conscious states that can be accessed over a short time is exceedingly large. For example, even if we just consider visual images, we can easily discriminate among innumerable scenes within a fraction of a second (9). More generally, the occurrence of a given conscious state implies an extremely rapid selection among a repertoire of possible conscious states that is, in fact, as large as one's experience and imagination. Differentiation among a repertoire of possibilities constitutes information, in the specific sense of reduction of uncertainty (10). Although this is often taken for granted, the occurrence of one particular conscious state over billions of others therefore constitutes a correspondingly large amount of information. Furthermore, it is information that makes a difference, in that it may lead to different consequences in terms of either thought or action.

The informativeness of consciousness helps dispose of many of the paradoxes raised about conscious experience. Consider a photodiode that can differentiate between light and dark and then provide an audible output, and a conscious human performing the same task and giving a verbal report. Why should the differentiation between light and dark performed by the human be associated with conscious experience, while presumably that performed by the photodiode is not? The paradox disappears if one considers the information generated by such discriminations. To the photodiode, the discrimination between darkness and light is the only one available, and is therefore minimally informative. To a conscious human, by contrast, an experience of complete darkness and an experience of complete light are two specific conscious experiences selected out of an enormous repertoire, and their selection implies the availability of a correspondingly large amount of information. To understand consciousness, it is important to identify

The authors are at The Neurosciences Institute, 10640 John J. Hopkins Drive, San Diego, CA 92121, USA.

\*To whom correspondence should be addressed. E-mail: tononi@nsi.edu

underlying neural processes that are both integrated and capable of such exceptionally informative differentiations.

### General Properties of Neural Processes Underlying Conscious Experience

Distributed neural activity, particularly in the thalamocortical system, is almost certainly essential for determining the contents of conscious experience (4, 11). We suggested previously that a key neural mechanism underlying conscious experience are the reentrant interactions between posterior thalamocortical areas involved in perceptual categorization and anterior areas related to memory, value, and planning for action. Such interactions among neuronal groups in distributed brain areas may be necessary in order to generate a unified neural process corresponding to a multimodal conscious scene (4). Recent experimental findings are consistent with this hypothesis and suggest some generalizations about the neural processes that underlie conscious experience.

*Activation and deactivation of distributed neuronal populations.* Changes in specific aspects of conscious experience correlate with changes in activity in specific brain areas, whether the experience is driven by external stimuli, by memory, or by imagery and dreams (12). Conscious experience as such, however, involves the activation or deactivation of widely distributed brain areas (13), although what should count as the appropriate reference state for comparison is not clear. In subjects who are comatose or deeply anesthetized, unconsciousness is associated with a profound depression of neural activity in both the cerebral cortex and thalamus (13). During slow-wave sleep, in which consciousness is severely reduced or lost, cerebral blood flow is globally reduced as compared to both waking and REM (rapid eye movement) sleep, two brain states associated with vivid conscious reports (14). A more specific reference state would be the response to a simple sensory input when a subject is unaware of it versus when the subject is aware of it. We have used magnetoencephalography to measure brain responses to flickering visual stimuli under conditions of binocular rivalry (15). A vertical grating flickering at one frequency was presented to one eye and a horizontal grating, flickering at a different frequency, was presented to the other eye. Although the stimuli were presented together, the subjects perceived either the vertical grating or the horizontal grating, with an alternation every few seconds. It was found that the power of steady-state neuromagnetic responses at the frequency of the flickering stimulus (its frequency tag) was higher by 30 to 60% in many sensor locations when the subject was con-

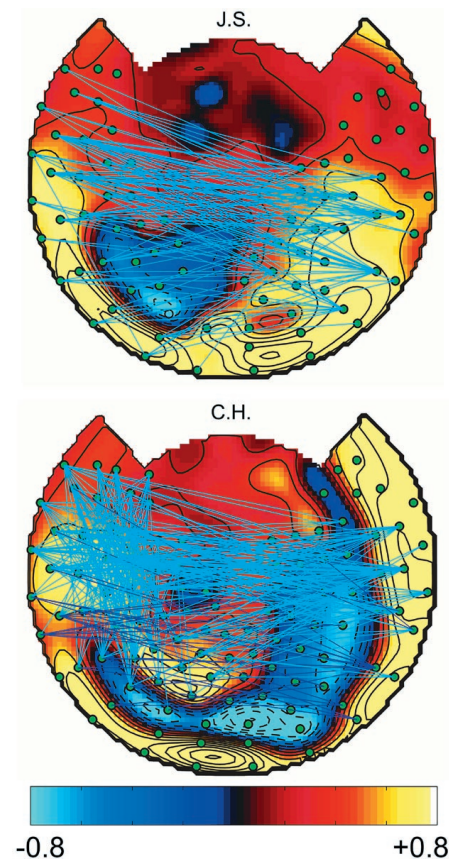
scious of that stimulus. The sensors with frequency tags that correlated with conscious experience were widely distributed over both posterior (occipital and temporal) and anterior (frontal) areas. Furthermore, there were considerable variations among different subjects (Fig. 1).

A change in the degree to which neural activity is distributed within the brain may accompany the transition between conscious, controlled performance and unconscious, automated performance. When tasks are novel, brain activation related to the task is widely distributed; when the task has become automatic, activation is more localized and may

shift to a different set of areas (16). In animal studies, neural activity related to sensory stimuli can be recorded in many brain regions before habituation. After habituation sets in (a time when humans report that stimuli tend to fade from consciousness), the same stimuli evoke neural activity exclusively along their specific sensory pathways (17). These observations suggest that when tasks are automatic and require little or no conscious control, the spread of signals that influence the performance of a task involves a more restricted and dedicated set of circuits that become "functionally insulated." This produces a gain in speed and precision, but a loss in context-sensitivity, accessibility, and flexibility (18).

*Integration through strong and rapid reentrant interactions.* Activation and deactivation of distributed neural populations in the thalamocortical system are not sufficient bases for conscious experience unless the activity of the neuronal groups involved is integrated rapidly and effectively. We have suggested that such rapid integration is achieved through the process of reentry—the ongoing, recursive, highly parallel signaling within and among brain areas. Large-scale computer simulations have shown that reentry can achieve the rapid integration or "binding" of distributed, functionally specialized neuronal groups dynamically, that is, in a unified neural process rather than in a single place (19, 20).

Substantial evidence indicates that the integration of distributed neuronal populations through reentrant interactions is required for conscious experience. An indication comes from the study of patients with disconnection syndromes, in which one or more brain areas are anatomically or functionally disconnected from the rest of the brain due to some pathological process (21). In the paradigmatic disconnection syndrome (the split brain), visual or somatosensory stimuli can activate the nondominant hemisphere and lead to behavioral responses, but the dominant, verbal hemisphere is not aware of them (22). Although the two hemispheres can still communicate through indirect, subcortical routes, rapid and effective neural interactions mediated by direct reentrant connections are abolished by the lesion of the corpus callosum. Modeling studies suggest that a telltale sign of effective reentrant interactions is the occurrence of short-term temporal correlations between the neuronal groups involved (19). Experiments on cats show that short-term temporal correlations between the activity of neuronal groups responding to the same stimulus, but located in different hemispheres, are abolished by callosal transections (23). Other studies indicate that various kinds of cognitive tasks are accompanied by the occurrence of short-term temporal correlations among



**Fig. 1.** Amplitude and coherence differences between the steady-state neuromagnetic responses during binocular rivalry when subjects were conscious of a stimulus and when they were not. The differences are taken between amplitude and coherence values at 7.41 Hz when the subjects were conscious of a vertical grating flickered at 7.41 Hz and when they were not (that is, when they were conscious of a horizontal grating flickered at 9.5 Hz). Amplitude differences are topographically displayed for two subjects. Color scale is in picotesla. Significant positive differences in coherence at 7.41 Hz between pairs of distant sensors are indicated by superimposed cyan lines. Blue lines indicate negative differences in coherence. Filled green circles indicate channels with signal-to-noise ratio  $> 5$  that have coherence values  $> 0.3$  with at least one other channel. See (15) for details.

distributed populations of neurons in the thalamocortical system (24). The magnetoencephalographic study of binocular rivalry mentioned above (15) also indicates that awareness of a stimulus is associated with increased coherence among distant brain regions (Fig. 1).

The requirement for fast, strong, and distributed neural interactions may explain why stimuli that are feeble, degraded, or short-lasting, often fail to be consciously perceived. Although such stimuli may produce a behavioral response [perception without awareness (25, 26)], they are unlikely to ignite neural activity of sufficient strength or duration to support fast distributed interactions. Conversely, attention may increase the conscious salience of certain stimuli by boosting the corresponding neural responses as well as the strength of neural interactions (27). Neural activity is also more likely to contribute effectively to distributed neural interactions if it is sustained for hundreds of milliseconds. This would lead to the functional closure of longer reentrant loops and thereby support reentrant interactions among more distant regions (19, 20). Experimental findings are consistent with this idea. High-frequency somatosensory stimuli delivered to the thalamus require about 500 ms for the production of a conscious sensory experience, while less than 150 ms are sufficient for sensory detection without awareness (28). The sustained evoked potentials associated with a conscious somatosensory sensation are apparently generated by the excitation of pyramidal neurons of primary somatosensory cortex through reentrant interactions with higher cortical areas (29).

Evidence for a correlation between conscious experience and sustained neural activity also comes from tasks involving visuospatial working memory—the ability to rehearse or “keep in mind” a spatial location. Working memory is used to bring or keep some item in consciousness or close to conscious access (30). In working memory tasks, sustained neural activity is found in prefrontal cortex of monkeys, and it is apparently maintained by reentrant interactions between frontal and parietal regions (31). Sustained neural activity may facilitate the integration of the activity of spatially segregated brain regions into a coherent, multimodal neural process that is stable enough to permit decision-making and planning (32).

*Differentiated patterns of activity.* Although strong and fast reentrant interactions among distributed groups of neurons are necessary for conscious experience, in themselves, they are still not sufficient. This is strikingly demonstrated by the unconsciousness accompanying generalized seizures and slow-wave sleep. During generalized seizures, the brain is not only extremely active,

but most neurons fire in a highly synchronous manner. For example, the electroencephalogram (EEG) during petit mal absences indicates that groups of neurons over the whole brain are either all firing together or all silent together, with these two neural states alternating every third of a second. Although such hypersynchronous firing is indicative of strong and distributed interactions, a subject who is prey to such a seizure is unconscious. Similarly, during slow-wave sleep, neurons in the thalamocortical system are active as well as remarkably interactive, as shown by their synchronous firing in a stereotyped, burst-pause pattern. During this stage of sleep, however, it is rare to obtain vivid and extensive conscious reports (33). By contrast, during REM sleep, when neural activity is not globally synchronous but resembles the rapid and complex patterns of waking, subjects typically report vivid dreams if awakened. We suggest that the low-voltage, fast-activity EEG characteristic of waking and REM sleep reflects the availability of a rich and diverse repertoire of neural activity patterns. If the repertoire of differentiated neural states is large, consciousness is possible. Conversely, if this repertoire is reduced, as when most groups of neurons in the cortex discharge synchronously and functional discriminations among them are obliterated, consciousness is curtailed or lost (34).

### Theoretical Concepts and Measures

This brief review of neurological and neurophysiological data indicates that the distributed neural process underlying conscious experience must be functionally integrated and at the same time highly differentiated. As mentioned above, two key properties of conscious experience are that it is integrated, in the sense that it cannot be subdivided into independent components, and that it is extremely differentiated, in the sense that it is possible, within a short time, to select among an enormous number of different conscious states. It is a central claim of this article that analyzing the convergence between these phenomenological and neural properties can yield valuable insights into the kinds of neural processes that can account for the corresponding properties of conscious experience. Such an analysis requires the availability of satisfactory measures of integration and differentiation that can be applied to actual neural processes, as well as an understanding of the neural mechanisms of integration.

*Functional clustering: How to identify an integrated process.* How can one determine whether a neural process is unified or simply a collection of independent or nearly independent subprocesses? We have suggested that a subset of distributed elements within a system gives rise to a single, integrated process if, at a given time scale, these elements

interact much more strongly among themselves than with the rest of the system — for example, if they form a functional cluster. This criterion has been formalized by introducing a direct measure of functional clustering (35) which we summarize here.

Consider a  $j$ th subset of  $k$  elements ( $X_j^k$ ) taken from an isolated neural system  $X$ , and its complement ( $X - X_j^k$ ). Interactions between the subset and the rest of the system introduce statistical dependence between the two. This is measured most generally by their mutual information  $MI(X_j^k; X - X_j^k) = H(X_j^k) + H(X - X_j^k) - H(X)$ , which captures the extent to which the entropy of  $X_j^k$  is accounted for by the entropy of  $X - X_j^k$  and vice versa [H indicates statistical entropy (36)]. The statistical dependence within a subset can be measured by a generalization of mutual information, which is called integration and is given by  $I(X_j^k) = \sum H(x_i) - H(X_j^k)$ , where  $H(x_i)$  is the entropy of each element  $x_i$  considered independently. We then define the functional cluster index  $CI(X_j^k) = I(X_j^k) / MI(X_j^k; X - X_j^k)$  as a ratio of the statistical dependence within the subset and the statistical dependence between that subset and the rest of the system. Based on this definition, a subset of neural elements that has a CI value much higher than 1 and does not itself contain any smaller subset with a higher CI value constitutes a functional cluster. This is a single, integrated neural process that cannot be decomposed into independent or nearly independent components.

We have applied these measures of functional clustering both to simulated datasets and to positron emission tomography data obtained from schizophrenic subjects performing cognitive tasks (35). Theoretically sound measures that can detect the occurrence of functional clustering at the time scale (fractions of a second) crucial for conscious experience may require additional assumptions. Nevertheless, it would appear that the rapid establishment of synchronous firing among cortical regions and between cortex and thalamus should be considered as an indirect indicator of functional clustering, since it implies strong and fast neural interactions among the neural populations involved (19, 20). The mechanisms of rapid functional clustering among distributed populations of neurons in the thalamocortical system have been studied with the help of large-scale simulations (19, 20). These have shown that the emergence of high-frequency synchronous firing in the thalamocortical system depends critically on the dynamics of corticothalamic and corticocortical reentrant circuits and on the opening of voltage-dependent channels in the horizontal corticocortical connections (37).

*Neural complexity: Measuring the differences that make a difference.* Once an inte-

grated neural process is identified, we need to determine to what degree that process is differentiated. Does it give rise to a large repertoire of different activity patterns or neural states? It is essential to consider only those differences between activity patterns that make a difference to the system itself. A TV screen may, for example, go through a large number of "activity patterns" that look different to an external observer, but that make no difference to the TV.

A possible approach to measuring differences that make a difference within an integrated neural system is to consider it as its own "observer." This can be achieved by dividing the system (which, we assume, constitutes a functional cluster) into two subsets and then measuring their mutual information (38). The value of  $MI(X_j^k; X - X_j^k)$  between a  $j$ th subset  $X_j^k$  of the isolated system  $X$  and its complement  $X - X_j^k$  will be high if two conditions are met. Both  $X_j^k$  and  $X - X_j^k$  must have many states [their entropy must be relatively high (10)], and the states of  $X_j^k$  and of  $X - X_j^k$  must be statistically dependent (the entropy of  $X_j^k$  must be largely accounted for by the interactions with  $X - X_j^k$ , and vice versa). The expression  $MI(X_j^k; X - X_j^k)$  reflects how much, on average, changes in the state of  $X - X_j^k$  make a difference to the state of  $X_j^k$ , and vice versa.

To obtain an overall measure of how differentiated a system is, one can consider not just a single subset of its constituent elements, but all its possible subsets. The corresponding measure, called neural complexity, is given by  $C_N(X) = \frac{1}{2} \sum (MI(X_j^k; X - X_j^k))$ , where the sum is taken over all  $k$  subset sizes and the average is taken over all  $j$ th combinations of  $k$  elements. Complexity is thus a function of the average mutual information between each subset and the rest of the system, and it reflects the number of states of a system that result from interactions among its elements (39).

It can be shown that high values of complexity reflect the coexistence of a high degree of functional specialization and functional integration within a system, as appears to be the case for systems such as the brain. For example, the dynamic behavior of a simulated cortical area containing thousands of spontaneously active neuronal groups (38) resembled the low-voltage fast-activity EEG of waking states and had high complexity. Such a system, whose connections were organized according to the rules found in the cortex, visited a large repertoire of different activity patterns that were the result of interactions among its elements. If the density of the connections was reduced, the dynamic behavior of the model resembled that of a noisy TV screen and had minimal complexity. A large number of activity patterns were visited, but they were merely the result of the

independent fluctuations of its elements. If the connections within the cortical area were instead distributed at random, the system yielded a hypersynchronous EEG that resembled the high-voltage waves of slow-wave sleep or of generalized epilepsy. The system visited a very limited repertoire of activity patterns, and its complexity was low.

Measures of complexity, like measures of functional clustering, can also be applied to neurophysiological data to evaluate the degree to which a neural process is both integrated and differentiated (40). This opens the way to comparisons of the values of neural complexity in different cognitive and arousal states and to empirical tests of the relationships between brain complexity and conscious experience.

### The Dynamic Core Hypothesis

A final issue we should consider is whether the neural process underlying conscious experience extends to most of the brain, as was concluded by William James, or is restricted to varying subsets of neuronal groups. Several observations support the latter possibility.

1) Classical lesion and stimulation studies suggest that many brain structures outside the thalamocortical system have no direct influence on conscious experience. Even within the thalamocortical system, many regions can be lesioned or stimulated without producing direct effects on conscious experience (41).

2) Neurophysiological studies indicate a possible dissociation between conscious experience and ongoing neural activity within portions of the thalamocortical system. During binocular rivalry in monkeys, a large proportion of neurons in early visual areas, such as V1, V4, and MT, continued to fire to their preferred stimulus even when it was not consciously perceived (42). The activity of only a subset of the neurons recorded in these areas was correlated with the percept, although in higher areas such as IT and STS, the percentage reached 95%. In our magnetoencephalographic study of binocular rivalry in humans (Fig. 1) (15), we found that the responses of only a subset of occipital, temporal, and frontal areas was correlated with the conscious perception of a stimulus, although several other regions showed widespread responses to stimuli that were not consciously perceived.

3) The firing of neurons dealing with rapidly varying local details of a sensory input or a motor output does not seem to map to conscious experience. The latter deals with invariant properties of objects that are highly informative as well as more stable and easily manipulated. For example, patterns of neural activity in the retina and other early visual structures correspond faithfully to spatial and temporal details of the visual input and are in constant flux. During each visual fixation,

however, humans extract the meaning of a scene and are not conscious of considerable changes in its local details (43). Groups of neurons responding in a stable way to invariant properties of objects are therefore more likely to contribute to conscious experience.

4) Many neural processes devoted to carrying out highly automated routines that make it possible to talk, listen, read, write, and so forth, in a fast and effortless way do not appear to contribute directly to conscious experience, although they are essential in determining its content (44). As mentioned above, neural circuits carrying out such highly practiced neural routines may become functionally insulated except at the input or output stages. There is also some evidence that cortical regions that are part of a fast system for controlling action, such as the dorsal visual stream, may not contribute significantly to conscious experience (45).

5) Although the sheer anatomical connectivity of the brain may hint that, over a sufficiently long time scale, everything can interact with everything else, modeling studies indicate that only certain interactions within the thalamocortical system are fast and strong enough to lead to the formation of a large functional cluster within a few hundred milliseconds (46).

These observations suggest that changes in the firing of only certain distributed subsets of the neuronal groups that are activated or deactivated in response to a given task are associated with conscious experience. What is special about these subsets of neuronal groups, and how can they be identified? We suggest the following:

1) A group of neurons can contribute directly to conscious experience only if it is part of a distributed functional cluster that achieves high integration in hundreds of milliseconds.

2) To sustain conscious experience, it is essential that this functional cluster be highly differentiated, as indicated by high values of complexity.

We propose that a large cluster of neuronal groups that together constitute, on a time scale of hundreds of milliseconds, a unified neural process of high complexity be termed the "dynamic core," in order to emphasize both its integration and its constantly changing activity patterns. The dynamic core is a functional cluster: its participating neuronal groups are much more strongly interactive among themselves than with the rest of the brain. The dynamic core must also have high complexity: its global activity patterns must be selected within less than a second out of a very large repertoire.

The dynamic core would typically include posterior corticothalamic regions involved in perceptual categorization interacting reentrantly with anterior regions involved in concept for-

mation, value-related memory, and planning (4), although it would not necessarily be restricted to the thalamocortical system. The term "dynamic core" deliberately does not refer to a unique, invariant set of brain areas (be they prefrontal, extrastriate, or striate cortex), and the core may change in composition over time (47). Because our hypothesis highlights the role of the functional interactions among distributed groups of neurons rather than their local properties (2), the same group of neurons may at times be part of the dynamic core and underlie conscious experience, while at other times it may not be part of it and thus be involved in unconscious processes. Furthermore, since participation in the dynamic core depends on the rapidly shifting functional connectivity among groups of neurons rather than on anatomical proximity, the composition of the core can transcend traditional anatomical boundaries (48). Finally, as suggested by imaging studies (15), the exact composition of the core related to particular conscious states is expected to vary significantly across individuals.

The dynamic core hypothesis avoids the category error of assuming that certain local, intrinsic properties of neurons have, in some mysterious way, a privileged correlation with consciousness. Instead, this hypothesis accounts for fundamental properties of conscious experience by linking them to global properties of particular neural processes. We have seen that conscious experience is a process that is unified and private, that is extremely differentiated, and that evolves on a time scale of hundreds of milliseconds. The dynamic core is a process, since it is characterized in terms of time-varying neural interactions, not as a thing or a location. It is unified and private, because its integration must be high at the same time as its mutual information with what surrounds is low, thus creating a functional boundary between what is part of it and what is not. The requirement for high complexity means that the dynamic core must be highly differentiated—it must be able to select, based on its intrinsic interactions, among a large repertoire of different activity patterns. Finally, the selection among integrated states must be achieved within hundreds of milliseconds, thus reflecting the time course of conscious experience (49).

A number of experimental questions and associated predictions are generated by this hypothesis. A central prediction is that, during cognitive activities involving consciousness, there should be evidence for a large but distinct set of distributed neuronal groups that interact over fractions of a second much more strongly among themselves than with the rest of the brain. This prediction could, in principle, be tested by recording, in parallel, multiple neurons whose activity is correlated with conscious experience. Multielectrode recordings have already indicated that rapid

changes in the functional connectivity among distributed populations of neurons can occur independently of firing rate (50). Recent studies in monkey frontal cortex also show abrupt and simultaneous shifts among stationary activity states involving several, but not all recorded neurons (51). A convincing demonstration of rapid functional clustering among distributed neuronal groups requires, however, that these studies be extended to larger populations of neurons in several brain areas. Another possibility would be to examine whether the effects of direct cortical microstimulation spread more widely in the brain if they are associated with conscious experience than if they are not. In humans, the extent and boundaries of neural populations exchanging coherent signals can be evaluated through methods of frequency tagging (15). Techniques offering both wide spatial coverage and high temporal resolution could also help establish how large a dynamic core normally is, how its composition changes, and whether certain brain regions are always included or always excluded. It is also significant to ask whether the dynamic core can split, and thus whether multiple dynamic cores can coexist in a normal subject. A reasonable prediction would be that certain disorders of consciousness, notably dissociative disorders and schizophrenia, should be reflected in abnormalities of the dynamic core and possibly result in the formation of multiple cores.

A strong prediction based on our hypothesis is that the complexity of the dynamic core should correlate with the conscious state of the subject. For example, we predict that neural complexity should be much higher during waking and REM sleep than during the deep stages of slow-wave sleep, and that it should be extremely low during epileptic seizures despite the overall increase in brain activity. We also predict that neural processes underlying automatic behaviors, no matter how sophisticated, should have lower complexity than neural processes underlying consciously controlled behaviors. Finally, a systematic increase in the complexity of coherent neural processes is expected to accompany cognitive development.

The outcome of such tests should indicate whether conscious phenomenology can indeed be related, as we suggest, to a distributed neural process that is both highly integrated and highly differentiated. The evidence available so far supports the belief that a scientific explanation of consciousness is becoming increasingly feasible (52).

#### References and Notes

1. W. James, *The Principles of Psychology* (Holt, New York, 1890).
2. F. Crick and C. Koch, *Cold Spring Harbor Symp. Quant. Biol.* **55**, 953 (1990); *Nature* **375**, 121 (1995);

- S. Zeki and A. Bartels, *Proc. R. Soc. London Ser. B* **265**, 1583 (1998).
3. G. Ryle, *The Concept of Mind* (Hutchinson, London, 1949).
4. G. M. Edelman, *The Remembered Present* (Basic Books, New York, 1989); ——— and G. Tononi, *Consciousness: How Matter Becomes Imagination* (Basic Books, New York, in press); see also G. Tononi and G. M. Edelman, in *Consciousness*, H. Jasper et al., Eds. (Plenum, New York, 1998), pp. 245–280.
5. A "conscious state" is meant here as an idealization, exemplified by viewing a rapid succession of slides.
6. J. D. Holtzman and M. S. Gazzaniga, *Neuropsychologia* **23**, 315 (1985).
7. H. Pashler, *Psychol. Bull.* **116**, 220 (1994). The duration of this interval is comparable with the duration of conscious states [A. L. Blumenthal, *The Process of Cognition* (Prentice-Hall, Englewood Cliffs, NJ, 1977)].
8. F. Sengpiel, *Curr. Biol.* **7**, R447 (1997).
9. H. Intraub, *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 604 (1981); I. Biederman, *Science* **177**, 77 (1972).
10. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana, IL, 1963). Note that the informativeness of consciousness also helps us to understand its evolutionary value (4).
11. V. B. Mountcastle, in *The Mindful Brain*, G. M. Edelman and V. B. Mountcastle, Eds. (MIT Press, Cambridge, MA, 1978), p. 7; A. Damasio, *Cognition* **33**, 25 (1989); R. Llinas, U. Ribary, M. Joliot, X.-J. Wang, in *Temporal Coding in the Brain*, G. Buzsaki, R. Llinas, W. Singer, Eds. (Springer-Verlag, Berlin, 1994); J. Newman, *Consciousness Cognit.* **4**, 172 (1995); T. W. Picton and D. T. Stuss, *Curr. Biol.* **4**, 256 (1994).
12. R. S. J. Frackowiak, *Human Brain Function* (Academic Press, San Diego, CA, 1997); P. E. Roland, *Brain Activation* (Wiley-Liss, New York, 1993); M. I. Posner and M. E. Raichle, *Images of Mind* (Scientific American Library, New York, 1994). These imaging studies confirm and extend previous lesion and stimulation studies.
13. Lesion studies indicate that consciousness is abolished by widely distributed damage but not by localized cortical damage. The only localized brain lesions resulting in loss of consciousness typically affect the reticular core in the upper brainstem and hypothalamus or its rostral extensions in the reticular and intralaminar thalamic nuclei [F. Plum, in *Normal and Altered States of Function*, A. Peters and E. G. Jones, Eds. (Plenum, New York, 1991), vol. 9, p. 359]. Although it has been suggested that the reticular core may have a privileged connection to conscious experience [J. E. Bogen, *Consciousness Cognit.* **4**, 52 (1995)], its activity may simply be required to sustain distributed activity patterns in the cortex.
14. A. R. Braun et al., *Science* **279**, 91 (1998); P. Maquet et al., *Nature* **383**, 163 (1996). Neural activity in slow-wave sleep is reduced in both anterior neocortical regions (most of the prefrontal cortex), as well as in posterior cortical regions (especially parietal association areas), in paralimbic structures (anterior cingulate cortex and anterior insula), and in centrencephalic structures (reticular activating system, thalamus, and basal ganglia); in contrast, it is not depressed in unimodal sensory areas (primary visual, auditory, and somatosensory cortex).
15. G. Tononi, R. Srinivasan, D. P. Russell, G. M. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3198 (1998); R. Srinivasan, D. P. Russell, G. M. Edelman, G. Tononi, *Soc. Neurosci. Abstr.* **24**, 433 (1998).
16. S. E. Petersen, H. vanMier, J. A. Fiez, M. E. Raichle, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 853 (1998); R. J. Haier et al., *Brain Res.* **570**, 134 (1992).
17. J. A. Horel et al., *Science* **158**, 394 (1967).
18. B. J. Baars, *A Cognitive Theory of Consciousness* (Cambridge Univ. Press, New York, 1988).
19. O. Sporns, G. Tononi, G. M. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 129 (1991); G. Tononi, O. Sporns, G. M. Edelman, *Cereb. Cortex* **2**, 310 (1992).
20. E. D. Lumer, G. M. Edelman, G. Tononi, *Cereb. Cortex* **7**, 207 (1997); *ibid.*, p. 228. For example, in a large-scale model of the visual system, reentrant interactions between groups of neurons in perceptual or "posterior" areas and in executive or "anterior" areas

- rapidly led to their synchronous firing and to a correct behavioral discrimination. This discrimination was based on the dynamic binding of multiple visual attributes (position, movement, color, form) and of different levels of stimulus generalization (local features, invariant aspects of stimuli).
21. B. Kolb and I. Q. Whishaw, *Fundamentals of Human Neuropsychology* (Freeman, New York, 1996). Psychiatric dissociation syndromes and conversion disorders may originate from a similar alteration of reentrant interactions, although in these cases, the disconnection would be functional rather than anatomical [J. F. Kihlstrom, *Consciousness Cognit.* **1**, 47 (1992)]. Some explicit-implicit dissociations, such as amnesia, may also be due to a partial disconnection of a lesioned area from the more global pattern of neural activity that is associated with consciousness [D. L. Schacter, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11113 (1992)].
  22. M. S. Gazzaniga, *Neuron* **14**, 217 (1995).
  23. A. K. Engel, P. König, A. K. Kreiter, W. Singer, *Science* **252**, 1177 (1991).
  24. S. L. Bressler, *Brain Res. Rev.* **20**, 288 (1995); W. Singer and C. M. Gray, *Annu. Rev. Neurosci.* **18**, 555 (1995); M. Joliot, U. Ribary, R. Llinas, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11748 (1994); A. Gevins et al., *Electroencephalogr. Clin. Neurophysiol.* **98**, 327 (1996).
  25. A. J. Marcel, *Cognit. Psychol.* **15**, 238 (1983); *ibid.*, p. 197; P. M. Merikle, *Am. Psychol.* **47**, 792 (1992). In some cases, perception without awareness has been shown to occur with stimuli that are not short-lasting or weak [F. C. Kolb and J. Braun, *Nature* **377**, 336 (1995); S. He, H. S. Smallman, D. I. A. MacLeod, *Invest. Ophthalmol. Visual Sci.* **36**, S438 (1995).
  26. S. He, P. Cavanagh, J. Intriligator, *Nature* **383**, 334 (1996)].
  27. J. H. Maunsell, *Science* **270**, 764 (1995); K. J. Friston, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 796 (1998).
  28. B. Libet, *Ciba Found. Symp.* **174**, 123 (1993).
  29. L. Cauller, *Behav. Brain Res.* **71**, 163 (1995).
  30. A. Baddeley, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13468 (1996).
  31. J. M. Fuster, R. H. Bauer, J. P. Jervey, *Brain Res.* **330**, 299 (1985); P. S. Goldman-Rakic and M. Chafee, *Soc. Neurosci. Abstr.* **20**, 808 (1994).
  32. The idea that neural activity must persist for a minimum period of time in order to contribute to conscious experience is also suggested by the phenomenon of masking [J. L. Taylor and D. I. McCloskey, *Exp. Brain Res.* **110**, 62 (1996); K. J. Meador et al., *Neurology* **51**, 721 (1998)].
  33. M. Steriade, *Cereb. Cortex* **7**, 583 (1997); D. Kahn, E. F. Pace-Schott, J. A. Hobson, *Neuroscience* **78**, 13 (1997).
  34. Neural activity must also exhibit sufficient variance in time to support conscious perception. For example, if images on the retina are stabilized, perception fades rapidly, and a similar effect is seen in Ganzfeld stimulation. Short-lasting visual stimuli become invisible if the transient neuronal responses associated with their onset and offset are suppressed by masking stimuli [S. L. Macknik and M. S. Livingstone, *Nature Neurosci.* **1**, 144 (1998)].
  35. G. Tononi, A. R. McIntosh, D. P. Russell, G. M. Edelman, *Neuroimage* **7**, 133 (1998).
  36. As a measure of statistical dependence, mutual information has the virtue of being highly general, because it is multivariate and sensitive to high-order moments of statistical dependence [A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1991)]. Note that mutual information reflects a statistical dependence among subsets of a system, irrespective of its source. The presence and direction of causal interactions between two subsets of a system can be evaluated, at least in principle, by measuring the change in mutual information obtained by perturbing or deafferenting each subset in turn.
  37. These observations are of interest in view of the well-known action of certain so-called dissociative anesthetics, such as ketamine and phencyclidine, that act as noncompetitive antagonists of the *N*-methyl-D-aspartate receptor [H. Flohr, *Behav. Brain Res.* **71**, 157 (1995)].
  38. G. Tononi, O. Sporns, G. M. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5033 (1994). A complexity measure that does not involve the calculation of average values of integration and mutual information can also be defined as the amount of the entropy of a system that is accounted for by the interactions among its elements and is given by  $\sum MI(X_i^1; X - X_i^1) - I(X)$  [G. Tononi, G. M. Edelman, O. Sporns, *Trends Cognit. Sci.*, in press]. Note that complexity measures should be applied to a single system (a functional cluster) and not to a collection of independent or nearly independent subsystems.
  39. Changes in complexity can be obtained without modifying the anatomical connectivity of the model by simulating the transition between the burst-pause pattern of firing typical of slow-wave sleep and the tonic mode of firing typical of waking and REM sleep (G. Tononi, unpublished material). It should be noted that high complexity is not easy to achieve. A system of elements that are randomly interconnected, for instance, may look very complicated, but it has low values of complexity. On the other hand, systems that undergo selective processes so as to match the statistical structure of a rich environment will gradually increase their complexity [G. Tononi, O. Sporns, G. M. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3422 (1996)].
  40. K. J. Friston, G. Tononi, O. Sporns, G. M. Edelman, *Hum. Brain Mapp.* **3**, 302 (1995).
  41. W. Penfield, *The Excitable Cortex in Conscious Man* (Thomas, Springfield, IL, 1958).
  42. D. A. Leopold and N. K. Logothetis, *Nature* **379**, 549 (1996); D. L. Shenberg and N. K. Logothetis, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3408 (1997). For other instances of dissociation, see (26); M. Gur and D. M. Snodderly, *Vision Res.* **37**, 377 (1997); I. N. Pigarev, H. C. Nothdurft, S. Kastner, *Neuroreport* **8**, 2557 (1997); D. C. Bradley, G. C. Chang, R. A. Andersen, *Nature* **392**, 714 (1998).
  43. D. J. Simons and D. T. Levin, *Trends Cogn. Sci.* **1**, 261 (1997). The neurological evidence is in agreement with these psychological observations. In the adult, lesions of the retina produce blindness, but they do not eliminate visual imagery, visual memories, and visual dreams, while the latter are eliminated by lesions of certain visual cortical areas [M. Solms, *The Neuropsychology of Dreams* (Erlbaum, Mahwah, NJ, 1997)]. V1 may be important, however, to provide visual consciousness with a certain degree of detail. See also R. Jackendoff [*Consciousness and the Computational Mind* (MIT Press, Cambridge, MA, 1987)].
  44. R. M. Shiffrin, in *Scientific Approaches to Consciousness*, J. D. Cohen and J. W. Schooler, Eds. (Erlbaum, Mahwah, NJ, 1997), p. 49; L. L. Jacoby, D. Ste-Marie, J. P. Toth, in *Attention: Selection, Awareness, and Control*, A. D. Baddeley and L. Weiskrantz, Eds. (Clarendon, Oxford, 1993), p. 261; W. Schneider, M. Pimm-Smith, M. Worden, *Curr. Opin. Neurobiol.* **4**, 177 (1994).
  45. A. D. Milner, *Neuropsychologia* **33**, 1117 (1995); \_\_\_\_\_ and M. A. Goodale, *The Visual Brain in Action* (Oxford Univ. Press, New York, 1995).
  46. The organization of the anatomical connectivity of certain brain regions, such as the thalamocortical system, is much more effective in generating coherent dynamic states than that of other regions, such as the cerebellum or the basal ganglia (G. Tononi, unpublished material). Consistent with this, although in cortical and thalamic areas 20 to 50% of all pairs of neurons recorded are broadly synchronized, neurons in the internal segment of the globus pallidus, the output station of the basal ganglia, are almost completely uncorrelated [H. Bergman et al., *Trends Neurosci.* **21**, 32 (1998)].
  47. If the fast integration of neural activity comes at a premium in terms of number of connections and energetic requirements, neuronal groups in "higher" areas should be privileged members of the dynamic core underlying consciousness. Everything else being equal, their firing is more informative, in the sense that it rules out a larger number of possibilities. For example, the firing of face-selective neurons in area IT considerably reduces uncertainty about a visual scene (seeing a face rules out countless other visual scenes), while the firing of retinal neurons reduces uncertainty by much less (a bright spot in a certain position of the visual field is consistent with countless visual scenes). The results of studies of binocular rivalry in monkeys and humans mentioned above are consistent with this view.
  48. We emphasize that the dynamic core, the highly complex, rapidly established functional cluster proposed to underlie conscious experience, is in no way the only integrated but distributed neural process that is relevant to brain function. We have hypothesized that distributed but integrated neural processes called global mappings, encompassing portions of the thalamocortical system, as well as parallel loops through cortical appendages such as the basal ganglia, the hippocampus, and the cerebellum, underlie the unity of behavioral sequences (4). The functional integration of global mappings is envisioned to occur at longer time scales than the dynamic core (seconds as opposed to fractions of a second). However, these two kinds of dynamic processes are expected to partially overlap for short periods of time.
  49. Qualia—the seemingly inexplicable phenomenological manifestations of conscious experience—are conceived within this framework as rapid, highly informative discriminations within a repertoire of billions of neural states available to a unified neural process of great complexity. They correspond to the generation of a large amount of information in a short period of time. In this view, each quale—even a seemingly simple quale like a feeling of "redness"—corresponds to a discriminable state of the dynamic core in its entirety, and not merely to the state of a specific group of neurons in a certain brain area. The subjective meaning or quale of "redness," for example, would be defined by the (increased) activity of red-selective neurons as much as by the (reduced or unmodified) activity of neuronal groups selective for green or blue, for visual motion or shape, for auditory or somatosensory events, and for proprioceptive inputs, body schemas, emotions, intentions, and so forth, that jointly constitute the dynamic core. This view is antithetical to modular or atomistic approaches to consciousness (2).
  50. E. Vaadia et al., *Nature* **373**, 515 (1995).
  51. E. Seidemann, I. Meilijson, M. Abeles, H. Bergman, E. Vaadia, *J. Neurosci.* **16**, 752 (1996).
  52. It is perhaps worth pointing out that our analysis predicts the possibility of constructing a conscious artifact and outlines some key principles that should constrain its construction. This work was carried out as part of the theoretical neurobiology program at The Neurosciences Institute, which is supported by Neurosciences Research Foundation. The Foundation receives major support for this program from Novartis Pharmaceuticals Corporation and the W. M. Keck Foundation.