Consciousness and complexity

Anil K. Seth[a] **and Gerald M. Edelman**[b]

[a]Department of Informatics, University of Sussex, Brighton BN1 9QJ, UK
[b]The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA

E-mail: a.k.seth@sussex.ac.uk, www.anilseth.com

Article outline

Glossary

- *Thalamocortical system.* The network of highly interconnected cortical areas and thalamic nuclei that comprises a large part of the mammalian brain. The cortex is the wrinkled surface of the brain, the thalamus is a small walnut-sized structure at its center. An intact thalamocortical system is essential for normal conscious experience.
- *Theory of neuronal group selection (TNGS).* A large-scale selectionist theory of brain development and function with roots in evolutionary theory and immunology. According to this theory, brain dynamics shape and are shaped by selection among highly variant neuronal populations guided by value or salience.
- *Neural correlate of consciousness.* Patterns of activity in brain regions or groups of neurons that have privileged status in the generation of conscious experience.

Explanatory correlates are neural correlates that in addition account for key properties of consciousness.

- *Dynamic core.* A distributed and continually shifting coalesence of patterns of activity among neuronal groups within the thalamocortical system. According to the TNGS, neural dynamics within the core are of high neural complexity by virtue of which they give rise to conscious discriminations.
- *Neural complexity.* A measure of simultaneous functional segregation and functional integration based on information theory. A system will have high neural complexity if each of its components can take on many different states and if these states make a difference to the rest of the system.
- *Small-world networks.* Networks in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops or steps. Small-world networks combine high clustering with short path lengths. They can be readily identified in neuroanatomical data, and they are well suited to generating dynamics of high neural complexity.
- *Metastability.* Dynamics that are characterized by segregating and integrating influences in the temporal domain; metastable systems are neither totally stable nor totally unstable.

1. Definition of the subject and its importance

How do conscious experiences, subjectivity, and apparent free will arise from their biological substrates? In the mid 1600s Descartes formulated this question in a form that has persisted ever since [1]. According to Cartesian dualism consciousness exists in a non-physical mode, raising the difficult question of its relation to physical interactions in the brain, body and environment. Even in the late twentieth century, consciousness was considered by many to be outside the reach of natural science [2], to require strange new physics [3], or even to be beyond human analysis altogether [4]. Over the last decade however, there has been heightened interest in attacking the problem of consciousness through scientific investigation [5, 6, 7, 8, 9]. Succeeding in this inquiry stands as a key challenge for twenty-first century science.

Conventional approaches to the neurobiology of consciousness have emphasized the search for so-called 'neural correlates': Activity within brain regions or groups of neurons that has privileged status in the generation of conscious experience [10]. An important outcome of this line of research has been that consciousness is closely tied to neural activity in the thalamocortical system, a network of cortical areas and subcortical nuclei that forms a large part of the vertebrate brain [11, 12]. Yet correlations by themselves cannot supply explanations, they can only constrain them. A promising avenue toward explanation is to focus on key properties of conscious experience and to identify neural processes that can account for these properties; we can call these processes *explanatory correlates*. This article clarifies some of the issues surrounding

this approach and describes ways of characterizing quantitatively the *complexity* of neural dynamics as a candidate explanatory correlate.

Complexity is a central concept within many branches of systems science and more generally across physics, statistics, and biology; many quantitative measures have been proposed and new candidates appear frequently [13, 14, 15]. The complexity measures described in this article are distinguished by focusing on the extent to which a system's dynamics are *differentiated* while at the same time *integrated*. This conception of complexity accounts for a fundamental feature of consciousness, namely that every conscious experience is composed of many different distinguishable components (differentiation) and that every conscious experience is a unified whole (integration). According to the theoretical perspective described here, the combination of these features endows consciousness with a discriminatory capability unmatched by any other natural or artificial mechanism.

While the present focus is on consciousness and its underlying mechanisms, it is likely that the measures of complexity we describe will find application not only in neuroscience but also in a wide variety of natural and artificial systems.

2. Introduction

2.1. Consciousness

Consciousness is that which is lost when we fall into a dreamless sleep and returns when we wake up again. As William James emphasized, consciousness is a *process* and not a 'thing' [5]. Conscious experiences have content such as colors, shapes, smells, thoughts, emotions, inner speech, and the like, and are commonly accompanied by a sense of self and a subjective perspective on the world (the 'I'). The phenomenal aspects of conscious content (the 'redness' of red, the 'warmth' of heat, etc.) are in philosophical terminology called *qualia* [16].

It is important to distinguish between conscious *level*, which is a position on a scale from brain-death and coma at one end to vivid wakefulness at the other, and conscious *content*, which refers to composition of a conscious scene at a given (non-zero) conscious level. Obviously, conscious level and conscious content are related inasmuch as the range of possible conscious contents increases with conscious level (see Figure 1). It is also possible to differentiate *primary* (sensory) consciousness from *higher-order* (meta) consciousness [6, 17]. Primary consciousness refers to the presence of perceptual conscious content (colors, shapes, odors, etc.). Higher-order consciousness (HOC) refers to the fact that we are usually conscious of being conscious; that is, human conscious contents can refer to ongoing primary conscious experiences. HOC is usually associated with language and an explicit sense of selfhood [18] and good arguments can be made that primary consciousness can exist in principle in the absence of HOC, and that in many animals it probably does [19, 20].

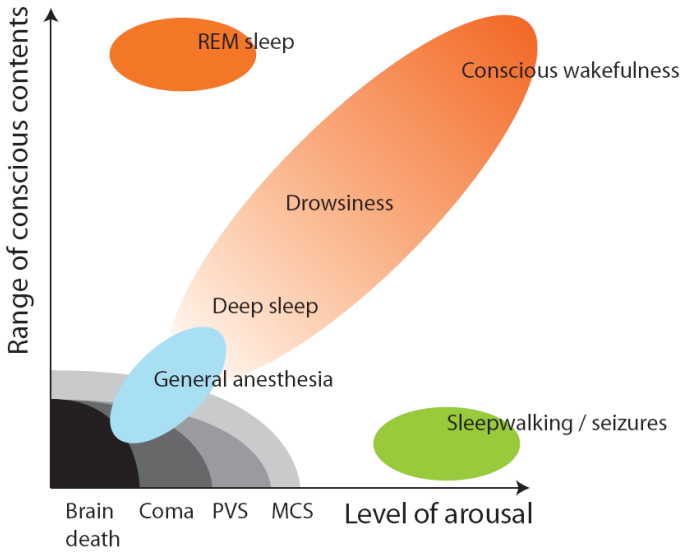

Figure 1. Conscious level is correlated with the range of possible conscious contents. PVS = persistent vegetative state, MCS = minimally conscious state. Adapted from [21].

2.2. Consciousness as discrimination

There are many aspects of consciousness that require explanation (see Table 1). However, one especially salient aspect that has been too often overlooked is that every conscious scene is both *integrated* and *differentiated* [22]. That is, every conscious scene is experienced 'all of a piece', as unified, yet every conscious scene is also composed of many different parts and is therefore one among a vast repertoire of possible experiences: When you have a particular experience, you are distinguishing it from an enormous number of alternative possibilities. On this view, conscious scenes reflect informative discriminations in a very high dimensional space where the dimensions reflect all the various modalities that comprise a conscious experience: sounds, smells, body signals, thoughts, emotions, and so forth (Figure 2).

Because the above point is fundamental, it is useful to work through a simple example (adapted from [22, 24]). Consider a blank rectangle that is alternately light and dark (Figure 3A,B). Imagine that this rectangle is all there is, that you are seated in front of it, and that you have been instructed to say "light" and "dark" as appropriate. A simple light-sensitive diode is also in front of the screen and beeps whenever the screen is light. Both you and the diode can perform the task easily, therefore both you and the diode can discriminate these two states: lightness and darkness. But each time the diode beeps, it is entering into one of a total of two possible states. It is minimally differentiated. However, when you say "light" or "dark" you are reporting one out of an enormous number of possible experiences. This point is emphasized by considering a detailed image such as a photograph (Figure 3C). A conscious person will

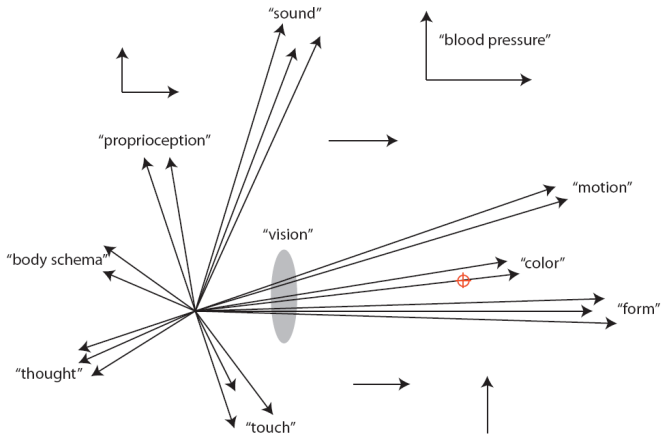

Figure 2. The figure shows an N-dimensional neural space corresponding to the dynamic core (see Section 3). N is the number of neuronal groups that, at any time, are part of the core, where N is normally very large (much larger than is plotted). The appropriate neural reference space for the conscious experience of 'pure red' would correspond to a discriminable point in the space (marked by the red cross). Focal cortical damage can delete specific dimensions from this space.

readily see this image as distinct both from the blank rectangle and from a scrambled version of the same image (Figure 3D). The diode, however, would classify both images and the rectangle as "light" (depending on its threshold), because it is insufficiently differentiated to capture the differences between the three.

Consider now an idealized digital camera. The electronics inside such a camera will enter a different state for the scrambled image than for the non-scrambled image; indeed, there will be a distinct state for any particular image. A digital camera is capable of much greater differentiation than the diode, but it is still not capable of discrimination because it is minimally *integrated*. In idealized form it is a collection of many independent light-sensitive diodes that must, to a good approximation, remain functionally independent from each other. From the perspective of this camera the image and the scrambled image are equivalent. *We* (as conscious organisms) can tell the difference between the two is because we integrate the many different parts of the image to form a coherent whole. We perceive each part of the image in relation to all the other parts, and we perceive each image in relation to all other possible images and possible conscious experiences that we may have. Successful discrimination therefore requires *both* integration *and* differentiation, and it can be hypothesized that it is this balance that yields the unity and diversity that is conscious experience.

Experimental evidence as well as intuition testifies to the fundamental nature of integration and differentiation in consciousness. A striking example is provided by so-called 'split brain' patients whose cortical hemispheres have been surgically separated. When presented with two independent visuospatial memory tasks, one

Table 1. Thirteen features of consciousness that require theoretical explanation. Items 1-6 are to some degree open to quantitative measurement whereas items 7-13 are more readily understood through logical and qualitative analysis. This list is drawn from [23] and a related list appears in [17].

1	Consciousness is accompanied by irregular, low-amplitude, fast (12-70 Hz) electrical brain activity.
2	Consciousness is associated with activity within the thalamocortical complex, modulated by activity in subcortical areas.
3	Consciousness involves distributed cortical activity related to conscious contents.
4	Conscious scenes are unitary.
5	Conscious scenes occur serially - only one conscious scene is experienced at a time.
6	Conscious scenes are metastable and reflect rapidly adaptive discriminations in perception and memory.
7	Conscious scenes comprise a wide multimodal range of contents and involve multimodal sensory binding.
8	Conscious scenes have a focus/fringe structure; focal conscious contents are modulated by attention.
9	Consciousness is subjective and private, and is often attributed to an experiencing 'self'.
10	Conscious experience is reportable by humans, verbally and non-verbally.
11	Consciousness accompanies various forms of learning. Even implicit learning initially requires consciousness of stimuli from which regularities are unconsciously extracted.
12	Conscious scenes have an allocentric character. They show intentionality, yet are shaped by egocentric frameworks.
13	Consciousness is a necessary aspect of decision making and adaptive planning.

to each hemisphere, they perform both very well [25]. In contrast, normal subjects cannot avoid integrating the independent signals into a single conscious scene which yields a much harder problem, and performance is correspondingly worse. In general, normal subjects are unable to perform multiple tasks simultaneously if they both require conscious input and they cannot make more than one conscious decision within the so-called 'psychological refractory period', a short interval of a few hundred milliseconds [26].

A loss of differentiation can be associated with the impoverishment of conscious contents following brain trauma. In 'minimally conscious' or 'persistent vegetative' states the dynamical repertoire of the thalamocortical system is reduced to the extent that adaptive behavioral responses are excluded [21]. In less dramatic cases focal cortical lesions can delete specific conscious contents; for example, damage to cortical region V4 can remove color dimensions from the space of possible experiences (cerebral achromatopsia [27]; c.f., Figure 2). Reportable conscious experience is also eliminated during generalized epileptic seizures and slow-wave sleep. Neural activity in these states is again poorly differentiated, showing hypersynchrony (epilepsy) or a characteristic synchronous 'burst pause' pattern (sleep) [22].

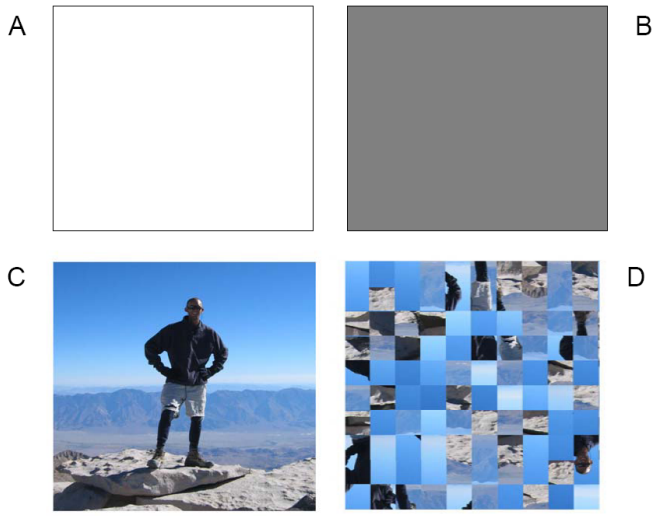

Figure 3. **A.** A light-colored rectangle. **B.** A dark-colored rectangle. **C.** A detailed image (the summit of Mount Whitney, California). **D.** A scrambled version of the same image. A simple light-sensitive diode would be able to discriminate **A** from **B**, but not among **A**, **C**, and **D**, since all these images would appear as 'light'. An idealized digital camera would enter a different state for each image **A**, **B**, **C**, and **D**, but would not discriminate between **C** and **D** because the camera does not integrate the various parts of each image to form a coherent whole. We can discriminate among all images because (i) our brain is capable of sufficient differentiation to enter a distinct state for each image, and (ii) our brain is capable of integrating the various parts of each image to form a coherent whole.

3. Consciousness and complexity

3.1. The dynamic core hypothesis

The notion that consciousness arises from neural dynamics that are simultaneously differentiated and integrated is expressed by the *dynamic core hypothesis* (DCH). This hypothesis has two parts [22, 28]:

- A group of neurons can contribute directly to conscious experience only if it is part of a distributed functional cluster (the dynamic core) that, through reentrant interactions in the thalamocortical system, achieves high integration in hundreds of milliseconds.
- To sustain conscious experience, it is essential that this functional cluster be highly differentiated, as indicated by high values of complexity.

The concept of a *functional cluster* refers to a subset of a neural system with dynamics that displays high statistical dependence internally and comparatively low statistical dependence with elements outside the subset: A functional cluster 'speaks mainly to itself' [29]. Conceiving of the dynamic core as a functional cluster implies that

the boundaries of the neural substrates of consciousness are continually shifting, with neuronal groups exiting and entering the core according to the flow of conscious contents and the corresponding discriminations being made. *Reentry* refers to the recursive exchange of signals among neural areas across massively parallel reciprocal connections and which in the context of the DCH serve to bind the core together. It is important to distinguish reentry from 'feedback' which refers to the recycling of an error signal from an output to an input [30, 31]. The interpretation of *complexity* in the context of the DCH is the subject of Section 4; for now we remark that it provides a quantitative measure of neural dynamics that is maximized by simultaneous high differentiation and high integration.

3.2. The theory of neuronal group selection

The DCH emerged from the theoretical framework provided by the 'theory of neural group selection' (TNGS), otherwise known as 'neural Darwinism' [32, 33, 18]. This section summarizes some of this essential background.

The TNGS is a biological perspective on brain processes with roots in evolutionary theory and immunology. It suggests that brain development and dynamics are *selectionist* in nature, and not instructionist, in contrast to computers which carry out explicit symbolic instructions. Four aspects of selectionist processes are emphasized: diversity, amplification/reproduction, selection, and degeneracy. *Diversity* in the brain is reflected in highly variant populations of neuronal groups where each group consists of hundreds to thousands of neurons of various types. This variation arises as a result of developmental and epigenetic processes such as cell division, migration, and axonal growth; subsequent strengthening and weakening of connections among cells (synapses) via experience and behavior generates further diversity. *Amplification* and *selection* in the brain are constrained by *value*, which reflects the salience of an event and which can be positive or negative as determined by evolution and learning. Value is mediated by diffuse ascending neural pathways originating, for example, in dopaminergic, catecholaminergic, and cholinergic brainstem nuclei [34]. As a result of value-dependent synaptic plasticity, connections among neuronal groups that support adaptive outcomes are strengthened, and those that do not are weakened. Finally, *degeneracy* emphasizes that in adaptive neural systems many structurally different combinations can perform the same function and yield the same output. Degeneracy is a key feature of many biological systems that endows them with adaptive flexibility [35, 36]. It is conspicuously absent in artificial systems which are correspondingly fragile (some artificial systems make use of 'redundancy' which differs from degeneracy in that specific functional units are explicitly duplicated; redundancy provides the robustness but not the flexibility of degeneracy).

According to the TNGS, primary consciousness arises when brain areas involved in ongoing perception are linked via reentry to brain areas responsible for a value-based memory of previous perceptual categorizations. On this view, primary consciousness

manifests as a 'remembered present' (akin to William James' 'specious present') by which an animal is able to exploit adaptive links between immediate or imagined circumstances and that animal's previous history of value driven behavior (figure 4).

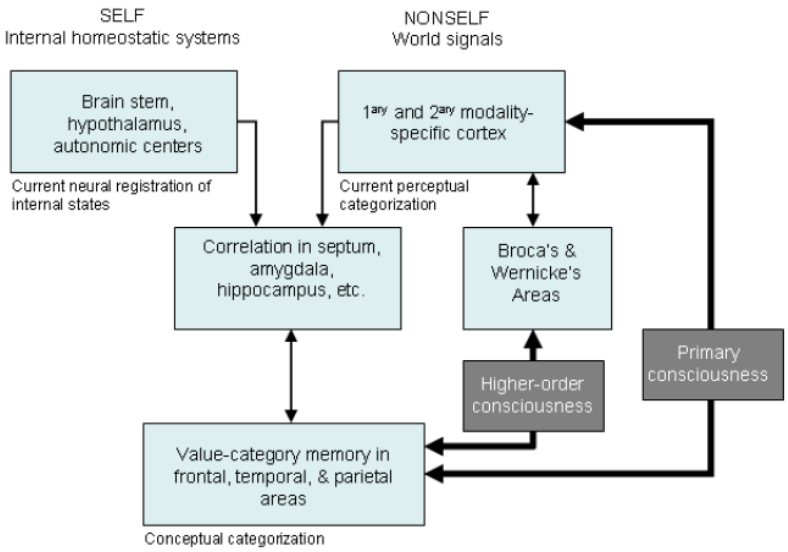

Figure 4. Primary consciousness and HOC in the TNGS. Signals related to value and signals from the world are correlated and produce value-category memories. These memories are linked by reentry to current perceptual categorization, resulting in primary consciousness. Higher-order consciousness depends on further reentry between value-category memory and current categorization via areas involved in language production and comprehension. Reprinted from [17].

The TNGS and the dynamic core hypothesis are closely related [18, 17]. They share the general claim that the neural mechanisms underlying consciousness arose in evolution for their ability to support multimodal discriminations in a high-dimensional space. In addition, the reentrant interactions linking immediate perception to value-category memory are precisely those that are suggested to bind together the dynamic core. Finally, the vast diversity of neural groups is central both to the original TNGS in providing a substrate for selection and to the DCH, in providing an essential component of neural complexity.

3.3. Consciousness and the dynamic core

We can now summarize the DCH and its origin in the TNGS. Consciousness is entailed by extensive reentrant interactions among neuronal populations in the thalamocortical system, the so-called dynamic core. These interactions, which support high-dimensional discriminations among states of the dynamic core, confer selective advantages on the organisms possessing them by linking current perceptual categorizations to value-

dependent memory. The high dimensionality of these discriminations is proposed to be a direct consequence of the rich complexity of the participating neural repertoires. Just as conscious scenes are both differentiated and integrated at the phenomenal level to yield high-dimensional discriminations, so too are the reentrant dynamics of their underlying neural mechanisms differentiated and integrated. Critically according to the TNGS, conscious qualia *are* the high-dimensional discriminations entailed by this balance of differentiation and integration as reflected in high complexity.

Any theory of consciousness must confront the question of whether conscious experiences have causal effects in the physical world [37]. Responding positively reflects common sense but it seems contrary to science to suggest non-physical causes for physical events. Responding negatively respects the causal closure of the physical world but appears to suggest that conscious experiences are 'epiphenomenal' and could in principle be done without (an implication that may be particularly troubling for experiences of 'free will' [38]). The TNGS addresses this quandary via the notion of *entailment*. According to the TNGS, dynamic core processes entail particular conscious experiences in the same way that the molecular structure of hemoglobin entails its particular spectroscopic properties: it simply could not be otherwise [18]. Therefore, although consciousness does not cause physical events, there exist particular physical causal chains (the neural mechanisms underlying consciousness) that by necessity entail corresponding conscious experiences: The conscious experience cannot be 'done without'.

3.4. Measuring consciousness and complexity

Having covered basic elements of the DCH and its origin in the TNGS, we turn now to the issue of measuring complexity in neural dynamics. To be useful in this context, candidate measures should satisfy several constraints. We have already mentioned that a suitable measure should reflect the fact that consciousness is a dynamic process [5], not a thing or a capacity. This point is particularly important in light of the observation that conscious scenes arise ultimately from transactions between organisms and environments, and these transactions are fundamentally processes [39]. (This characterization does not, however, exclude 'off-line' conscious scenes, for example those experienced during dreaming, reverie, abstract thought, planning, or imagery). A suitable measure should also take account of causal interactions within a neural system, and between a neural system and its surroundings - i.e., bodies and environments. Finally, to be of practical use, a suitable measure should also be computable for systems composed of large numbers of neuronal elements.

Obviously, the quantitative characterization of complexity can constitute only one aspect of a scientific theory of consciousness. This is true at both the neural level and at the level of phenomenal experience. At the neural level, no single measure could adequately describe the complexity of the underlying brain system (this would be akin, for example, to claiming that the complex state of the economy could be described by the gross domestic product alone). At the phenomenal level, conscious scenes have

many diverse features [18, 19], several of which do not appear to be readily quantifiable (see Table 1). These include subjectivity, the attribution of conscious experience to a self, and intentionality, which reflects the observation that consciousness is largely about events and objects. A critical issue nevertheless remains: how can measurable aspects of the neural underpinnings of consciousness be characterized?

4. Neural complexity

A fundamental intuition about complexity is that a complex system is neither fully ordered (e.g., a crystal) nor fully disordered (e.g., an ideal gas). This intuition is compatible with the central theme of the DCH, namely that the neural dynamics within the dynamic core should be both integrated and differentiated. The following definition of *neural complexity* (C_N), first proposed in 1994 [40], satisfies these intuitions and provides a practical means for assessing the complexity of neural and other systems.

4.1. Mathematical definition

Consider a neural system X composed of N elements (these may be neurons, neuronal groups, brain regions, etc.). A useful description of the dynamical connectivity of X is given by the joint probability distribution of the activities of its elements. Assuming that this function is Gaussian, this is equivalent to the covariance matrix of the system's dynamics COV(X). Importantly, COV(X) captures the total effect of all (structural) connections within a system upon deviation from statistical independence of the activities of a pair of elements, and not just the effect of any direct anatomical connection linking them [41]. Given COV(X) and assuming that the dynamics of X are covariance stationary (i.e., having unchanging mean and variance over time) the entropy of the system $H(\mathrm{X})$ is given by:

$$H(\mathrm{X}) = 0.5\ln((2\pi e)^N|\mathrm{COV(X)}|)$$

where $|.|$ denotes the matrix determinant [42]. $H(\mathrm{X})$ measures the overall degree of statistical independence exhibited by the system; i.e., its degree of differentiation. Knowing the entropy of a system allows calculation of the mutual information (MI) between two systems, or between two subsets of a single system. The MI between systems (or subsets) A and B measures the uncertainty about A that is accounted for by the state of B and is defined as $MI(\mathrm{A};\mathrm{B}) = H(\mathrm{A}) + H(\mathrm{B}) - H(\mathrm{AB})$ [43].

The integration of X, $I(\mathrm{X})$, measures the system's overall deviation from statistical independence. All elements in a highly integrated system are tightly coupled in their activity. With x_i denoting the i'th element of X, $I(\mathrm{X})$ can be calculated as:

$$I(\mathrm{X}) = \sum_{i=1}^{N} H(x_i) - H(\mathrm{X}).$$

$I(\mathrm{X})$ is equivalent to the measure 'multi-information' which was introduced several decades ago [44]. Having expressions for MI, $H(\mathrm{X})$, and $I(\mathrm{X})$ allows $C_N(\mathrm{X})$ to be

expressed in two equivalent ways. First, $C_N(\mathrm{X})$ can be calculated by summing the average MI between subsets of various sizes, for all possible bipartitions of the system:

$$C_N(\mathrm{X}) = \sum_k \langle MI(\mathrm{X}_j^k; \mathrm{X} - \mathrm{X}_j^k)\rangle, \tag{1}$$

where X_j^k is the j'th bipartition of size k, and $\langle . \rangle$ is the average across index j (figure 5A). $C_N(\mathrm{X})$ can also be expressed in terms of integration:

$$C_N(\mathrm{X}) = \sum_k \left((k/n)I(X) - \langle I(X_j^k)\rangle \right). \tag{2}$$

where $\langle I(X_j^k)\rangle$ is the average integration of all subsets of size k. $C_N(\mathrm{X})$ will be high if small subsets of the system show high statistical independence, but large subsets show low statistical independence. In other words, $C_N(\mathrm{X})$ will be high if each of its subsets can take on many different states and if these states make a difference to the rest of the system.

Because the full $C_N(\mathrm{X})$ can be computationally expensive to calculate for large systems, it is useful to have an approximation that considers only bipartitions consisting of a single element and the rest of the system. There are three mathematically equivalent ways of expressing this approximation, which is denoted $C(\mathrm{X})$:

$$\begin{aligned} C(\mathrm{X}) &= H(\mathrm{X}) - \sum_{k=1}^{N} H(x_i|\mathrm{X} - x_i) \\ &= \sum_i MI(x_i; X - x_i) - I(X) \\ &= (n-1)I(X) - n\langle I(X - x_i)\rangle, \end{aligned} \tag{3}$$

where $H(x_i|\mathrm{X} - x_i)$ denotes the conditional entropy of each element x_i given the entropy of the rest of the system $X - X_i$. These three expressions are equivalent for all X, whether they are linear or nonlinear, and neither $C_N(\mathrm{X})$ nor $C(\mathrm{X})$ can adopt negative values.

Recently, De Lucia *et al* [45] have developed a different approximation to $C_N(\mathrm{X})$ which is calculated directly from topological network properties (i.e., without needing covariance information). Their measure of 'topological $C_N(\mathrm{X})$' is based on the eigenvalue spectrum of the connectivity matrix of a network. While topological $C_N(\mathrm{X})$ offers substantial savings in computational expense it carries the assumption that the network is activated by independent Gaussian noise and therefore cannot be used to measure neural complexity in conditions in which a network is coupled to inputs and outputs (see Section 4.3 below).

4.2. Connectivity and complexity

There is a growing consensus that features of neuroanatomical organization impose important constraints on the functional dynamics underlying cognition [47, 48].

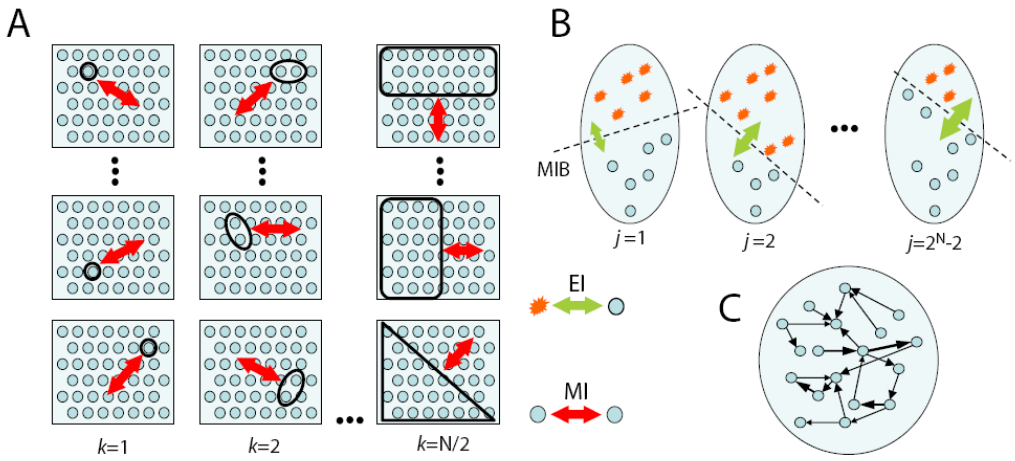

Figure 5. Measuring integration and differentiation in neural dynamics. **A.** Neural complexity C_N is calculated as the ensemble average mutual information (MI) between subsets of a given size and their complement, summed over all subset sizes (k) (adapted from Fig. 2 in [46]). Small circles represent neuronal elements and red arrows indicate MI between subsets and the remainder of the system. **B.** Information integration Φ is calculated as the effective information (EI) across the 'minimum information bipartition' (MIB). To calculate EI for a given bipartition (j), one subset is injected with maximally entropic activity (orange stars) and MI across the partition is measured. **C.** Causal density c_d is calculated as the fraction of interactions that are causally significant according to a mutlivariate Granger casuality analysis. A weighted (and unbounded) version of causal density (c_{dw}) can be calculated as the summed magnitudes of all significant causal interactions (depicted schematically by arrow width).

Accordingly, several studies have addressed the relationship between structural connectivity and neural complexity [49, 50, 51, 52, 53].

One useful approach employs evolutionary search procedures (genetic algorithms [54]) to specify the connection structure of simple networks under various fitness (cost) functions. A population of networks $\mathrm{X}_1 \ldots \mathrm{X}_N$ is initialized ('generation zero') with each member having random connectivity. Each network X_i is then evaluated according to a fitness function (for example, maximize $C(\mathrm{X})$) and those that score highly, as compared to the other networks in the population, are subjected to a small amount of random 'mutation' (i.e., small random changes in connectivity) and proceed to the next 'generation'. This procedure is repeated for many generations until the population contains networks that score near-optimally on the fitness function, or until the experimenter is satisfied that no further improvement is likely.

Sporns and colleagues applied a version of evolutionary search to find distinctive structural motifs associated with $H(\mathrm{X})$, $I(\mathrm{X})$, and $C(\mathrm{X})$ [49]. In this study, the initial population consisted of 10 networks each with $N = 32$ nodes and $K = 256$ connections and with fixed identical positive weights w_{ij}. The fitness function was determined by the

value of $H(\mathrm{X})$, $I(\mathrm{X})$, or $C(\mathrm{X})$ calculated from the covariance matrix of each network, assuming activation by covariance-stationary Gaussian noise. In each case they found that the resulting networks had distinctive structural features, as revealed both by simple visual inspection and by analysis using a variety of graph-theoretic measures. Networks optimized for $H(\mathrm{X})$ contained mostly reciprocal connections without any apparent local clustering. Networks optimized for $I(\mathrm{X})$ were highly clustered (i.e., neighboring nodes connect mainly to each other [55]) and had a long characteristic path length (i.e., a high mean separation between any two nodes in terms of number of intervening nodes). Finally, networks optimized for $C(\mathrm{X})$ had high clustering (and high reciprocal connectivity) coupled with a short characteristic path length. Strikingly, these networks were very similar to the so-called 'small world' class of network in which dense groups of nodes are connected by a relatively small number of reciprocal 'bridges' [55]. These networks also had a high proportion of 'cycles' (routes through the network that return to their starting point) and very low wiring lengths [49].

Sporns *et al* extended the above findings by calculating $C(\mathrm{X})$ for networks reflecting the known cortical connectivity of both the macaque visual cortex and the entire cat cortex. In both cases covariance matrices were obtained by assuming linear dynamics, equal connection strengths, and activation by covariance-stationary Gaussian noise. They found that both networks gave rise to high $C(\mathrm{X})$ as compared to random networks with equivalent distributions of nodes and connections. Indeed, the networks seemed to be near-optimal for $C(\mathrm{X})$ because random rewiring of connections led in almost all cases to a reduction in $C(\mathrm{X})$ [49].

In a separate study using a nonlinear neuronal network model including excitatory and inhibitory units, Sporns showed that regimes of high $C(\mathrm{X})$ coincided with 'mixed' connection patterns consisting of both local and long-range connections [56]. This result lines up with the previous study [49] in suggesting an association between small-world properties and complex dynamics. In addition, Sporns and Kötter found that networks optimized for the number of functional 'motifs' (small repeating patterns) had high $C(\mathrm{X})$ but those optimized for structural motifs did not [57] suggesting that high complexity reflects the presence of large functional repertoires. Finally, $C(\mathrm{X})$ seems to associate with fractal patterning, but not in a simple sense that fractal networks are optimal for complexity [58, 53]. Rather, fractality seems to be one among several structural attributes that contribute to the emergence of small-world features and complex dynamics. Together, these results indicate that only certain classes of network are able to support dynamics that combine functional integration with functional segregation and that these networks resemble in several ways those found in neuroanatomical systems.

4.3. Complexity and behavior

An important claim within the DCH is that complex dynamics provide adaptive advantages during behavior. To test this claim, Seth and Edelman examined examined

the relationship between behavior and neural complexity in a simple agent-based computational model [59]. They evolved networks similar to those in [49] (N=32, K=256) by selecting for their ability to guide target fixation behavior in a simulation model requiring coordination of 'head' and 'eye' movements (figure 6). Networks were evolved in both 'simple' and 'complex' environments where environmental complexity was reflected by unpredictable target movement and by variation in parameters affecting head and eye movement. Consistent with the DCH, networks supporting target fixation in rich environments showed higher C(X) than their counterparts adapted to simple environments. This was true both for dynamics exhibited during behavior in the corresponding environments ('interactive' complexity), and for dynamics evoked with Gaussian noise ('intrinsic' complexity).

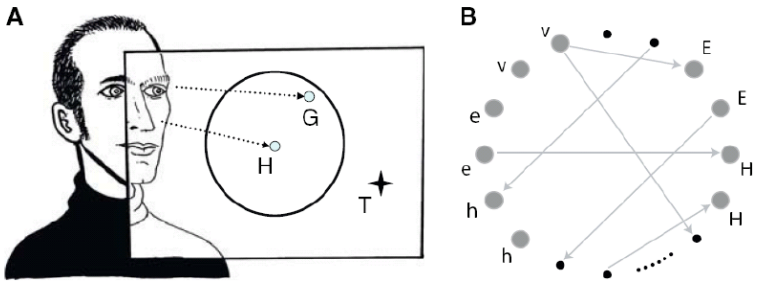

Figure 6. Target fixation model. **A.** The agent controls head-direction (H) and eye-direction (not shown) in order to move a gaze point (G) towards a target (T). **B.** Neural network controller. The 6 input neurons are shown on the left and the 4 output neurons on the right. Each pair of inputs (v,e,h) responds to x, y displacements: 'v' neurons to displacements of G from T, 'h' neurons to displacements of H from an arbitrary origin ('straight ahead'), and 'e' neurons to displacements of H from the eye-direction. The four output neurons control head direction (H) and eye-direction relative to the head (H). For clarity only 4 of the 22 interneurons are shown. Thin grey lines show synaptic connections. Only a subset of the 256 connections are shown. Adapted from [59].

Sporns and Lungarella explored the relationship between C(X) and behavior in a different way [60]. As in [59], networks acted as neural controllers during performance of a task (in this case, control of a simulated arm to reach for a target). However, instead of evolving for successful behavior, networks were evolved directly for high C(X). Strikingly, selecting for high C(X) led to networks that were able to perform the task, even though performance on the task had not been explicitly selected for. Finally, Lungarella and Sporns asked how C(X) depends on sensorimotor coupling by comparing neural dynamics of a robotic sensory array in two conditions: (i) unperturbed foveation behavior, and (ii) decoupling of sensory input and motor output via 'playing back' previously recorded motor activity [61]. They found significantly higher C(X) when sensorimotor coupling was maintained.

Taken together, the above results suggest a strong link between high neural

complexity and flexible, adaptive behavior. Of course, in none of these studies is any claim made that the corresponding networks are in any sense conscious.

4.4. Extensions and limitations

The concept of neural complexity has been extended to characterize the selectional responses of neural systems to inputs in terms of 'matching' complexity C_M [62]. C_M measures how well the intrinsic correlations within a neural system fit the statistical structure of a sensory stimulus. Simulations show that C_M is high when intrinsic connectivity is modified so as to differentially amplify those intrinsic correlations that are enhanced by sensory input, possibly reflecting the capacity of a neurally complex system to 'go beyond the information given' in a stimulus [62]. Despite this possibility C_M has not been investigated as thoroughly as has C_N.

C_N has several limitations. In its full form it is computationally prohibitive to calculate for large networks, but in approximation it is less satisfying as a measure. Also, C_N does not reflect complexity in the temporal domain since functional connections are analyzed at zero-lag [23]. Finally, C_N does not take into account *directed* (causal) dynamical interactions for the simple reason that MI is a symmetric measure. This last point is addressed by the alternative measures described below.

5. Information integration

The most prominent alternative to C_N is 'information integration' (Φ) [63, 24]. Unlike C_N, Φ reflects causal interactions because it is based on 'effective information' (EI), a directed version of MI that relies on the replacement of the outputs of different subsets of the studied system with maximum entropy signals.

5.1. Mathematical definition

Φ is defined as the effective information across the informational 'weakest-link' of a system, the so-called *minimum information bipartition* (MIB; figure 5B). It is calculated by the following procedure [63].

Given a system of N elements, identify all possible bipartitions of the system. For each bipartition $A|B$, replace the outputs from A by uncorrelated noise (i.e., maximally entropic activity), and measure how differentiated are the responses of its complement (B). This is the effective information (EI) between A and B:

$$\mathrm{EI}(A{\rightarrow}B) = \mathrm{MI}(A_{Hmax}; B),$$

where $\mathrm{MI}(A_{Hmax}; B)$ is the mutual information between A and B when the outputs from A have maximal entropy. $\mathrm{EI}(A{\rightarrow}B)$ measures the capacity for causal influence of partition A on its complement B (i.e., all possible effects of A on B). Given that $\mathrm{EI}(A{\rightarrow}B)$ and $\mathrm{EI}(B{\rightarrow}A)$ are not necessarily equal, one can define:

$$\mathrm{EI}(A{\leftrightarrow}B) = \mathrm{EI}(A{\rightarrow}B) + \mathrm{EI}(B{\rightarrow}A).$$

The minimum information bipartition (MIB) is the bipartition for which the normalized EI($A \leftrightarrow B$) is lowest. Normalization is accomplished by dividing EI($A \leftrightarrow B$) by $\min\{H_{max}(A); H_{max}(B)\}$, so that effective information is bounded by the maximum entropy available. The resulting MIB corresponds to the informational 'weakest link' of the system, and the Φ value of the system is the non-normalized EI($A \leftrightarrow B$) across the MIB.

A further stage of analysis has been described [63] in which a system can be decomposed into 'complexes' by calculating Φ for different subsets of elements; a complex is a subset having $\Phi > 0$ that is not included in a larger subset with higher Φ. For a given system, the complex with the maximum value of Φ is called the 'main complex'.

5.2. Information integration, connectivity, and complexity

As with neural complexity it is useful to explore what kinds of network structure lead to high values of Φ. Because of computational constraints only comparatively small networks have been investigated for their ability to generate high Φ (i.e., $N = 8$, $K = 16$ as opposed to $N = 32$, $K = 256$ as in [49]). In an initial study, networks optimized for Φ had highly heterogenous connectivity patterns with no two elements having the same sets of inputs and outputs [63]. At the same time, all nodes tended to emit and receive the same number of connections. These two properties arguably subserve functional segregation and integration respectively [63].

Although both Φ and C_N depend on a combination of functional integration and segregation, they are sensitive to different aspects of network dynamics. C_N reflects an average measure of integration that, unlike Φ, does not require heterogenous connectivity. On the other hand, unlike C_N, Φ is determined by the value of an informational meausure (EI) across only a single bipartition (the MIB) and is not modified by dynamical transactions across the remainder of the network. Finally, as mentioned above, Φ but not C_N is sensitive to causality.

5.3. Limitations and extensions

As with C_N, Φ does not measure complexity in the temporal domain [23]. There are also substantial limitations attending measurement of Φ for nontrivial systems. First, it is not possible in general to replace the outputs of arbitrary subsets of neural systems with uncorrelated noise. An alternative version of Φ can be envisaged in which 'transfer entropy' (TE) [64], a directed version of MI, is substituted for EI. TE can be calculated from the dynamics generated by a neural system during behavior and therefore does not require arbitrary perturbation of a system; it measures the *actual* causal influence across partitions whereas EI measures the *capacity* for causal influence. However, a version of Φ based on TE does not in general find the informational 'weakest link' (MIB) of a system since the MIB depends on capacity and not on transient dynamics.

Second, unlike C_N there is presently no well-defined approximation for Φ that removes the need to examine all possible bipartitions of a system. However, it may be

possible to make use of some informal heuristics. For example, bipartitions for which the normalized value of EI will be at a minimum will be most often those that cut the system in two halves, i.e., midpartitions [63]. Similarly, a representative rather than exhaustive number of perturbations may be sufficient to obtain at least an estimated value of Φ [63].

5.4. The information integration theory of consciousness

Φ occupies a central place in the 'information integration theory of consciousness' (IITC, [24]). According to this theory, consciousness *is* information integration as measured by Φ. The nature of the conscious content in a system with high Φ is determined by the particular informational relationships within the main complex (the complex with the highest Φ). While there are many similarities between the DCH and the IITC, most obviously that both make strong appeal to a measure of complexity, there are also important differences of which we emphasize two:

(i) Because Φ measures the capacity for information integration, it does not depend on neural activity *per se*. The IITC predicts that a brain where no neurons were active, but in which they were potentially able to react, would be conscious (perhaps of nothing). Similarly, a brain in which each neuron were stimulated to fire as an exact replica of your brain, but in which synaptic interactions had been blocked, would *not* be conscious [24]. The DCH has neither of these implications.

(ii) On the IITC Φ is an adequate measure of the 'quantity' of consciousness, therefore any system (biological or artificial) with sufficiently high Φ would necessarily be conscious. According to the DCH, high C_N is necessary but not sufficient for consciousness.

Point (ii) is particularly important in view of the finding that an arbitrarily high Φ can be obtained by a system as simple as a Hopfield network, which is a fully connected network with simple binary neuronal elements [23]. By choosing the synaptic strengths according to an exponential rule it can be shown that the corresponding Φ value scales linearly with network size, such that $\Phi(\mathrm{X}) = N$ bits for a network X of size N nodes. On the IITC this result leads to the counterintuitive conclusion that a sufficiently large Hopfield network will be conscious. Another challenge for the IITC in this context is the fact that the probability distributions determining entropy values (and therefore by extension Φ values) depend on subjective decisions regarding the spatial and temporal granularity with which the variables in a system are measured [[23, 65] but see [24]].

6. Causal density

A balance between dynamical integration and differentiation is likely to involve dense networks of causal interactions among neuronal elements. Causal density (c_d) is a measure of causal interactivity that captures both differentiated and integrated aspects

of these interactions [66, 23]. It differs from C_N by detecting causal interactions, differs from Φ by being sensitive to dynamical interactions across the whole network, and differs from both by being based not on information theory but instead on multivariate autoregressive modelling.

6.1. Mathematical definition

Causal density (c_d) measures the fraction of interactions among neuronal elements in a network that are causally significant (figure 5C). It can be calculated by applying 'Granger causality' [67, 68], a statistical concept of causality that is based on prediction: If a signal x_1 causes a signal x_2, then past values of x_1 should contain information that helps predict x_2 above and beyond the information contained in past values of x_2 alone [67]. In practice, Granger causality can be tested using multivariate regression modelling [69]. For example, suppose that the temporal dynamics of two time series, $x_1(t)$ and $x_2(t)$ (both of length T), can be described by a bivariate autoregressive model:

$$\begin{aligned} x_1(t) &= \sum_{j=1}^{p} A_{11,j} x_1(t-j) + \sum_{j=1}^{p} A_{12,j} x_2(t-j) + \xi_1(t) \\ x_2(t) &= \sum_{j=1}^{p} A_{21,j} x_1(t-j) + \sum_{j=1}^{p} A_{22,j} x_2(t-j) + \xi_2(t) \end{aligned} \tag{4}$$

where p is the maximum number of lagged observations included in the model (the model order, $p < T$), A contains the coefficients of the model, and ξ_1, ξ_2 are the residuals (prediction errors) for each time series. If the variance of ξ_1 (or ξ_2) is reduced by the inclusion of the x_2 (or x_1) terms in the first (or second) equation, then it is said that x_2 (or x_1) *G-causes* x_1 (or x_2). In other words, x_1 G-causes x_2 if the coefficients in A_{12} are jointly significantly different from zero. This relationship can be tested by performing an F-test on the null hypothesis that $A_{12,j} = 0$, given assumptions of covariance stationarity on x_1 and x_2. The magnitude of a significant interaction can be measured either by the logarithm of the F-statistic [70] or, more simply, by the log ratio of the prediction error variances for the restricted (R) and unrestricted (U) models:

$$\begin{aligned} gc_{2\rightarrow 1} &= log\frac{var(\xi_{1R(12)})}{var(\xi_{1U})} \text{ if } gc_{2\rightarrow 1} \text{ is significant,} \\ &= 0 \text{ otherwise,} \end{aligned}$$

where $\xi_{1R(12)}$ is derived from the model omitting the $A_{12,j}$ (for all j) coefficients in equation (4) and ξ_{1U} is derived from the full model.

Importantly, G-causality is easy to generalize to the multivariate case in which the G-causality of x_1 is tested in the context of multiple variables $x_2 \ldots x_N$. In this case, x_2 G-causes x_1 if knowing x_2 reduces the variance in x_1's prediction error when the activities of all other variables $x_3 \ldots x_n$ are also included in the regression model.

Both bivariate and multivariate G-causality have been usefully applied to characterizing causal interactions in simulated [71, 72] and biological [73] neural systems.

Following a Granger causality analysis, both normalized (c_d) and non-normalized (c_{dw}) versions of causal density of a network X with N nodes can be calculated as:

$$c_d(\mathrm{X}) = \frac{\alpha}{N(N-1)}, \qquad c_{dw}(\mathrm{X}) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} g c_{j \to i}$$

where α is the total number of significant causal interactions and $N(N-1)$. While normalized causal density is bounded to the range [0,1] the non-normalized version is unbounded.

High causal density indicates that elements within a system are both globally coordinated in their activity (in order to be useful for predicting each other's activity) and at the same time dynamically distinct (reflecting the fact that different elements contribute in different ways to these predictions). Therefore, as with both C_N and Φ, c_d reflects both functional integration and functional segregation in network dynamics.

6.2. Conditions leading to high causal density

In terms of connectivity, computational models show that both fully connected networks (having near-identical dynamics at each node) and a fully disconnected networks (having independent dynamics at each node) have low c_d and c_{dw}; by contrast, randomly connected networks have much higher values [71]. More detailed connectivity studies remain to be conducted.

An initial attempt to analyze behavioral conditions leading to high causal density was made by [66] revisiting the model of target fixation described previously [59]. To recapitulate, in this model networks were evolved in both 'simple' and 'complex' environments where environmental complexity was reflected by unpredictable target movement and by variation in parameters affecting head and eye movement (Figure 6). Causal density in this model was calculated from first-order differenced time series of the ten sensorimotor neurons and it was found that highest values of causal density occurred for networks evolved and tested in the 'complex' environments. These results mirrored those obtained with C_N, indicating an association between a high value of a complexity measure and adaptive behavior in a richly structured environment.

6.3. Extensions and limitations of causal density

A practical problem for calculating causal density is that multivariate regression models become difficult to estimate accurately as the number of variables (i.e., network elements) increases. For a network of N elements, the total number of parameters in the corresponding multivariate model grows as pN^2, and the number of parameters to be estimated for any single time series grows linearly (as pN), where p is the model order (equation 4). We note that these dependencies are much lower than the factorial dependency associated with Φ and C_N, and may therefore may be more readily

circumvented. One possible approach may involve the use of Bayesian methods for limiting the number of model parameters via the introduction of prior constraints on significant interactions [74]. In neural systems, such prior constraints may be derived, for example, on the basis of known neuroanatomy or by anatomically-based clustering procedures.

Several other extensions to causal density are suggested by enhancements to the statistical implementation of Granger causality:

- Nonlinear G-causality methods based, for example, on radial-basis-function kernels allow causal density to detect both linear and nonlinear causal interactions [75, 76].
- Partial G-causality (based on partial coherence) enhances robustness to common input from unobserved variables, supporting more accurate estimates of causal density in systems which cannot be fully observed [77].
- G-causality has a frequency-dependent interpretation [70, 73] allowing causal density to be assessed in specific frequency bands.

6.4. Causal density and consciousness

Although causal density is not attached to any particular theory of consciousness, it aligns closely with the DCH because it is inherently a measure of process rather than capacity. Causal density cannot be inferred from network anatomy alone, but must be calculated on the basis of explicit time series representing the dynamic activities of network elements during behavior. It also depends on all causal interactions within the system, and not just on those interactions across a single bipartition, as is the case for Φ. Finally, causal density incorporates the temporal dimension more naturally than is the case for either C_N or Φ; while the latter measure functional interactions at zero-lag only, causal density incorporates multiple time lags as determined by the order parameter p (equation 4).

The foregoing descriptions make clear that although existing formal measures may have heuristic value in identifying functional integration and functional segregation in neural dynamics, they remain inadequate in varying degrees. C_N can reflect process, can be computed for large systems in approximation, but does not capture causal interactions. Φ captures causal interactions, is infeasible to compute for large neural systems, and can be shown to grow without bound even for certain simple networks. Also, Φ is a measure of capacity rather than process but this is a deliberate feature of the IITC. c_d reflects all causal interactions within a system and is explicitly a measure of process, but it also is difficult to compute for large systems. An additional and important practical limitation of C_N, Φ, and c_d is that they apply only to statistically stationary dynamics.

7. Empirical evidence

We turn now to empirical evidence relevant to the DCH. Much of this evidence comes from patients with focal brain lesions and neuroimaging of healthy subjects using functional magnetic resonance imaging (fMRI), electroencephelography (EEG) and magnetoencephalography (MEG). Although current experimental methods are not sufficiently powerful to confirm or refute the DCH, their application, separately and in combination, yields much useful information. A detailed review appears in [17]; below we select some pertinent features.

7.1. Involvement of the thalamocortical system

A wealth of experimental evidence attests to thalamocortical involvement in consciousness, as demonstrated by both clinical studies and by experiments using normal subjects. Clinical studies show that damage to non-specific (intralaminar) thalamic nuclei can abolish consciousness *in toto* [78], whereas damage to cortical regions often deletes specific conscious features such as color vision, visual motion, conscious experiences of objects and faces, and the like [79]. No other brain structures show these distinctive effects when damaged.

Conscious functions in normal subjects are usefully studied by comparison with closely matched controls who perform the function unconsciously, an approach known as 'contrastive analysis' [80, 81, 82]. An emerging consensus among contrastive studies is that conscious contents correspond to widespread thalamocortical activation as compared to unconscious controls [81]. For example, Dehaene and colleagues have shown widespread fMRI activation peaks in parietal, prefrontal, and other cortical regions for conscious perception of visual words, as compared to unconscious inputs which activated mainly primary visual cortex [83]. Along similar lines, a recent fMRI study of motor sequence learning showed a shift from widespread cortical involvement during early learning (when conscious attention is required) to predominantly subcortical involvement during later learning phases (when skill production is comparatively 'automatic') [84].

7.2. Dynamical correlates: Binocular rivalry

A classical example of contrastive analysis makes use of the phenomenon of binocular rivalry, in which different images are projected to each eye [85]. Because of the integrative nature of consciousness these images, if sufficiently different, are not combined into a single composite; rather, conscious experience alternates between them. Srinivasan and colleagues used magnetoencephalography (MEG) to measure brain responses to flickering visual stimuli under rivalrous conditions [86, 87]. A vertical grating flickering at one frequency was presented to one eye and a horizontal grating flickering at another frequency was presented to the other; these different frequencies allowed stimulus-specific brain responses to be isolated in the neuromagnetic signal, a

technique known as 'frequency tagging' [88]. As expected for such stimuli, subjects perceived only one grating at any given time. It was found that the power of the frequency-tag of a stimulus was higher by 30-60% across much of the cortical surface when that stimulus was perceptually dominant compared to when it was perceptually suppressed. Moreover, there was a large increase in coherence among distant brain regions, consistent with the idea that conscious perception is associated with widespread integration of neural dynamics mediated by reentry. Although this coherence increase is not a direct test of the DCH it is consistent with the theory and underscores the value of looking for *dynamical* correlates of consciousness.

Cosmelli and colleagues have used a similar paradigm to show that development of a perceptual dominance period arises in neural terms as an extended dynamical process involving the propagation of activity throughout a distributed brain network beginning in occipital regions and extending into more frontal regions [89]. Such a 'wave of consciousness' might reflect underlying integrative processes that lead to the formation of a dynamic core [90]. Chen and colleagues modified the rivalry paradigm so that subjects saw both gratings with both eyes but had to differentially *pay attention* to one or the other [91]. Power increases but not coherence increases were found for the attended stimulus, suggesting that attention may not involve the same global integrative processes implicated in consciousness. Finally, Srinivasan has shown that coherence increases during dominance are due partly to increased phase locking to the external stimulus and partly to increased synchrony among intrinsic network elements, again in line with the idea that consciousness involves coalescence of a distributed functional cluster within the brain [92].

7.3. Sleeping, waking, and anesthesia

Binocular rivalry involves constant conscious level and changing conscious *content*. Experimental evidence relevant to conscious *level* comes from studies involving transitions between sleeping and waking, anesthesia, epileptic absence seizures and the like. Many studies have tracked changes in endogenous activity across these various transitions but direct assessments of specific complexity measures are mostly lacking. Nonetheless, current findings are broadly consistent with the DCH. As noted in Section 2.2, absence seizures and slow-wave sleep (but not rapid-eye-movement sleep) are characterized by hypersynchronous neural activity that may correspond to reduced functional segregation [22]. Anesthesia has particular promise for further experimental study because global anesthetic states can be induced via a wide variety of pharamacological agents having diverse physiological effects. Moreover, proposed unifying frameworks, such as Mashour's 'cognitive unbinding' theory [93], share with the DCH the idea that loss of consciousness can arise from diminished functional integration. In line with Mashour's proposal, John and colleagues have observed at anesthetic loss-of-consciousness (i) functional disconnection along the rostrocaudal (front-to-back) axis and across hemispheres (measured by coherence changes), and (ii) domination of the

EEG power spectrum by strongly anteriorized low frequencies [94].

The development of transcranial magnetic stimulation (TMS) has opened the possibility of studying effective (causal) connectivity in the brain. TMS noninvasively disrupts specific cortical regions by localized electromagnetic induction. Massimini and colleagues combined high-density EEG with TMS to test whether effective connectivity among distant brain regions is diminished during sleep [95]. Applying a TMS pulse to premotor cortex during quiet wakefulness led to a sequence of waves propagating throughout cortex, but this widespread propagation was mostly absent during non-rapid eye movement (slow wave) sleep. While these results do not allow calculation of any specific complexity measure they are consistent with both the IITC and the DCH.

In summary, it is clear that enhanced experimental and analytical methods are needed in order to test adequately whether C_N (or other specific measures) are modulated as predicted by the DCH. Initial attempts to calculate C_N directly from neural dynamics have not been successful (see [96] for a review) although a link to complexity is suggested by the discovery of 'small-world' networks in functional brain dynamics [97, 96].

8. Related theoretical proposals

8.1. Dynamical systems theory and metastability

We have already mentioned that the measures of complexity discussed in this article apply only to statistically stationary dynamics (Section 6.4). This restriction contrasts sharply with an alternative tradition in theoretical neuroscience which focuses on non-stationary brain dynamics and which emphasizes the tools of dynamical systems theory. This alternative tradition can be traced back to early suggestions of Turing [98] and Ashby [99] and was concisely expressed by Katchalsky in 1974: " ... waves, oscillations, macrostates emerging out of cooperative processes, sudden transitions, patterning, etc., seem made to order to assist in the understanding of integrative processes in the nervous system" [100]. More recently the dynamical systems approach has been championed in neuroscience by, among others, Haken [101] under the rubric 'coordination dynamics' and Freeman who has produced a steady stream of papers exploring dynamical principles in brain activity [102, 103]. Valuable reviews of work in this tradition can be found in [104, 105, 106].

A key concept in the dynamical systems approach is 'metastability' which describes dynamics that are "distinguished by a balanced interplay of integrating and segregating influences" [107] (p.26). While this definition is obviously similar to the intuition driving neural complexity, metastability has been fleshed out, not in the concepts of information theory or time-series analysis, but instead in the language of attractor dynamics. A dynamical system inhabits a metastable regime when there are no stable fixed points but only partial attraction to certain phase relationships among the system variables. At the level of neural dynamics metastability may reflect the ongoing creation and

dissolution of neuronal assemblies across distributed brain regions [107, 105, 108]. A now classical experimental example of metastability comes from a study in which subjects were asked to flex a finger in response to a periodic tone, initially in a syncopated manner [107]. As the frequency of the tone increases the syncopated response becomes harder to maintain until a critical point is reached at which the subject switches to a synchronous mode of response. Strikingly, this behavioral phase transition is accompanied by a corresponding transition in the patterning of neuromagnetic cortical signals. At the critical point, where there is partial attraction to both syncopated and synchronous response modes, both behavioral and neural dynamics are dominated by metastability. For other evidence of metastability in the brain see [109].

Metastability characterizes an important aspect of conscious experience, namely that conscious events are rapidly adaptive and fleeting [17]. Consciousness is remarkable for its present-centeredness [6, 5]. Immediate experience of the sensory world may last at most a few seconds and our fleeting cognitive present is surely less than half a minute in duration. This present-centeredness has adaptive value for an organism by allowing time enough to recruit a broad network of task-related neural resources while permitting neural dynamics to evolve responses to subsequent events. Thus, conscious experience can be described by an interplay of segregating and integrating influences in both the temporal (metastability) and spatial (complexity) domains. A key theoretical challenge is to work out in greater detail the relationship between these two concepts.

8.2. Global workspace theory

Beginning in 1988 [80] Baars has developed a cognitive theory of consciousness under the rubric 'global workspace (GW) theory' [80, 81, 110]. The cornerstone of GW theory is the idea that consciousness involves a central resource (the GW) which enables distribution of signals among numerous otherwise informationally encapsulated and functionally independent specialized processors. GW theory states that mental content becomes conscious mental content when it gains access to the GW such that it can influence a large part of the brain and a correspondingly wide range of behaviors. A key aspect of GW theory is that conscious contents unfold in an integrated, serial manner but are the product of massively parallel processing among the specialized processors. The integrated states of the GW follow each other in a meaningful but complex progression that depends on multiple separate processes, each of which might have something of value to add to the ongoing constitution of the GW. Although these notions are compatible the DCH, they do not by themselves specify dynamical properties to the same level of detail.

A dynamical approach to GW theory has been pursued by Wallace [112] and separately by Dehaene, Changeux and colleagues [113, 111, 114]. Wallace adopts a graph-theoretic perspective proposing that the GW emerges as a 'giant component' among transient collections of otherwise unconscious processors. The formation of a giant component in graph theory denotes a phase transition at which multiple sub-

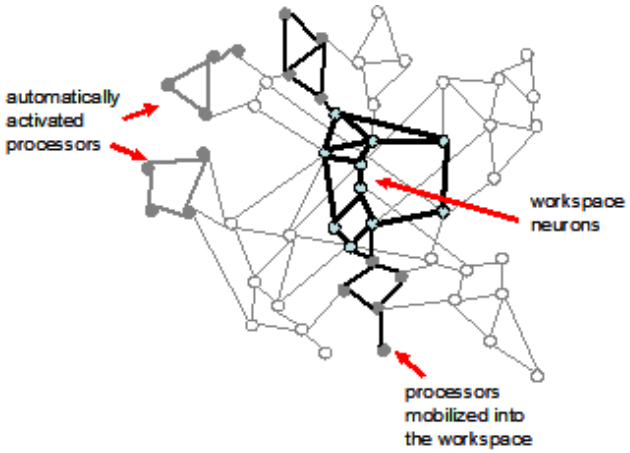

Figure 7. A schematic of the neuronal global workspace. A central global workspace, constituted by long-range cortico-cortical connections, assimilates other processes according to their salience. Other automatically activated processors do not enter the global workspace. Adapted from [111].

networks coalesce to form a large network including the majority of network nodes [115]. Dehaene and colleagues have built a series of computational models inspired by GW theory, to account for a variety of psychological phenomena including 'attentional blink' [111], 'inattentional blindness' [114], and effortful cognitive tasks [113]. These models are based on the concept of a 'neuronal global workspace' in which sensory stimuli mobilize excitatory neurons with long-range cortico-cortical axons, leading to the genesis of global activity patterns among so-called 'workspace neurons' (Figure 7). This model, and that of Wallace, predicts that consciousness is 'all or nothing' - i.e., a gradual increase in stimulus visibility should be accompanied by a sudden transition (ignition) of the neuronal GW into a corresponding activity pattern. As with Wallace, although some dynamic properties of the neuronal GW have been worked out and are compatible with the DCH, a rigorous account of how the model relates to neural complexity has not been attempted.

8.3. Neuronal synchrony and neuronal coalitions

The association of neural synchrony with consciousness arose from its proposed role as a mechanism for solving the so-called 'binding problem', which in general terms refers to problem of coordinating functionally segregated brain regions. The binding problem is most salient in visual perception for which the functional and anatomical segregation of visual cortex contrasts sharply with the unity of a visual scene. Since the 1980s a succession of authors have proposed that the binding problem is solved via neuronal synchronization [116, 117] and both experimental evidence [118, 119] and computational models have borne out the plausibility of this mechanism [31]. In the 1990s, starting with an early paper by Crick and Koch [7], this proposal grew into the

hypothesis that consciousness itself is generated by transient synchronization among widely distributed neuronal assemblies, with particular emphasis on oscillations in the gamma band (∼40 Hz) [120, 39]. In support of this idea we have already seen that conscious perception correlates with increased synchrony of a (non-gamma-band) visual 'frequency tag' [87], and several studies have reported associations between gamma-band synchrony and consciousness [121, 122, 123]. However, synchrony-based theories of binding (and by extension consciousness) remain controversial [124] and there is not yet evidence that disruptions of gamma-band synchrony lead to disruptions of conscious contents [12].

From the perspective of the DCH a deeper concern with the synchrony hypothesis is that it accounts only for integration, and not for the combination of integration and differentiation that yields the discriminatory power of consciousness. In a recent position paper [125], Crick and Koch reversed their previous support for gamma-band synchrony as a sufficient mechanism for consciousness, favoring instead the notion of competition among 'coalitions' of neurons in which winning coalitions determine the contents of consciousness at a given time. Such neuronal coalitions bear similarities to the decades-old notion of Hebbian assemblies [126] on a very large and dynamic scale. They also suggest that unconscious processing may consist largely in feed-forward processing whereas consciousness may involve standing waves created by bidirectional signal propagation, a proposal advanced as well by Lamme [127]. Crick and Koch note that the 'coalition' concept is similar to the dynamic core concept [125] although lacking in the detailed formal specification of the latter.

A possible role for gamma-band synchrony in both the DCH and in Crick and Koch's framework is that it may facilitate the formation but not the ongoing activity of the core (or a coalition) [125]. In this light it is suggestive that correlations between gamma-band synchrony and consciousness tend to occur at early stages of conscious perception [121, 122].

9. Outlook

Scientific accounts of consciousness continue to confront the so-called 'hard problem' of how subjective, phenomenal experiences can arise from 'mere' physical interactions in brains, bodies, and environments [128, 129]. It is possible that new concepts will be required to overcome this apparent conceptual gap [130]. It is equally likely that increasing knowledge of the mechanisms underlying consciousness will lead these philosophical conundrums to fade away, unless they have empirical consequences [81, 125]. In short, to expect a scientific resolution to the 'hard problem' as it is presently conceived may be to misunderstand the role of science in explaining nature. A scientific theory cannot presume to replicate the experience that it describes or explains; a theory of a hurricane is not a hurricane [18]. If the phenomenal aspect of experience is irreducible, so is the fact that physics has not explained why there is something rather than nothing, and this ontological limit has not prevented physics from laying bare many

mysteries of the universe.

The approach described in this article is one of developing *explanatory correlates* of consciousness, namely properties of neural dynamics that are experimentally testable and that account for key properties of conscious experience. Thanks to accelerating progress in experimental techniques and increased attention to theory, the outlook for this approach is healthy. We close by suggesting some key areas for further study:

- **Development of a large-scale model of a dynamic core**. Although progress in large scale neural network modelling has been rapid [131], we currently lack a sufficiently detailed model of environmentally-coupled thalamocortical interactions needed to test the mechanistic plausibility of the DCH. Having such a model should also allow substantive connections to be drawn between the DCH and GW theory.
- **Development of new experimental methods**. New methods are needed to track neuronal responses at sufficient spatio-temporal resolutions to support accurate estimation of C_N and other complexity measures during different conscious and unconscious conditions. Among current methods fMRI has poor time resolution and measures neural activity indirectly, while MEG/EEG lacks spatial acuity and is unable to record details of thalamic responses.
- **Complexity and metastability**. New theory is needed to relate the class of complexity measures described in this article to metastability, which analyzes functional segregation and integration in the temporal domain.
- **Emergence and 'downward causality'**. New theory is also needed to better understand how global dynamical states arise from their basal interactions and how these global states can constrain, enslave, or otherwise affect properties at the basal level [39]. Applied to consciousness and to cognitive states in general, such 'downward causality' can suggest functional roles and may even help reconcile the phenomenology of free-will with physiological fact.

10. Bibliography

10.1. Books and reviews

(i) Baars, D.J., Banks, W.P. and Newman, J.B., eds. (2003) Essential sources in the scientific study of consciousness. MIT Press, Cambridge, MA.

(ii) Dorogovtsev, S.N. and Mendes, J.F.F. (2003). Evolution of Networks: from biological networks to the Internet and WWW. Oxford University Press. Oxford, UK.

(iii) Edelman, G.M. Wider than the sky: The phenomenal gift of consciousness. Yale University Press, New Haven, CT.

(iv) Edelman, G.M. and Tononi, G. (2000). A universe of consciousness : How matter becomes imagination. Basic Books, New York, NY.

(v) Metzinger, T. ed. (2000) Neural correlates of consciousness. MIT Press, Cambridge, MA.

(vi) Koch, C. (2004). The Quest for Consciousness: A Neurobiological Approach, Roberts and co.

(vii) Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(1):42.

10.2. Primary literature

[1] E. Haldane and G. Ross, editors. *The philosophical work of Descartes*. Cambridge University Press, Cambridge, UK, 1985.

[2] K. Popper and J.F. Eccles. *The self and its brain*. Springer, New York, NY, 1977.

[3] R. Penrose. *Shadows of the mind: A search for the missing science of consciousness*. Oxford University Press, Oxford, UK, 1994.

[4] C McGinn. *The problem of consciousness*. Blackwell, Oxford, UK, 1991.

[5] W. James. Does consciousness exist? *J. Philos. Psychol. Sci. Methods*, 1:477–491, 1904.

[6] G.M. Edelman. *The remembered present*. Basic Books, Inc., New York, NY, 1989.

[7] F. Crick and C. Koch. Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2:263–275, 1990.

[8] T.C. Dalton and B.J. Baars. Consciousness regained: The scientific restoration of mind and brain. In T.C. Dalton and R.B. Evans, editors, *The lifecycle of psychological ideas: Understanding the prominence and the dynamics of intellectual change*, pages 203–247. Springer, Berlin, 2003.

[9] F. Crick and C. Koch. *The quest for consciousness: A neurobiological approach*. Roberts and co., 2004.

[10] T. Metzinger, editor. *Neural correlates of consciousness: Empirical and conceptual questions*. MIT Press, Cambridge, MA, 2000.

[11] R. Llinás, U. Ribary, D. Contreras, and C. Pedroarena. The neuronal basis for consciousness. *Philos Trans R Soc Lond B Biol Sci*, 353:1841–1849, 1998.

[12] G. Rees, G. Kreiman, and C. Koch. Neural correlates of consciousness in humans. *Nature Reviews Neuroscience*, 3(4):261–70, 2002.

[13] J.P Crutchfield and K. Young. Inferring statistical complexity. *Phys Rev Lett*, 63:105–108, 1989.

[14] W.H. Zurek, editor. *Complexity, entropy, and the physics of information*. Addison-Wesley, Redwood City, CA, 1990.

[15] C. Adami. What is complexity? *Bioessays*, 24:1085–1094, 2002.

[16] M. Tye. Qualia. In E. Zalta, editor, *The Stanford Encyclopedia of Philosophy (Fall 2007 Edition)*. 2007.

[17] A.K. Seth and B.J. Baars. Neural Darwinism and consciousness. *Consciousness and Cognition*, 14:140–168, 2005.

[18] G.M. Edelman. Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences, USA*, 100(9):5520–5524, 2003.

[19] A.K. Seth, B.J. Baars, and D.B. Edelman. Criteria for consciousness in humans and other mammals. *Consciousness and Cognition*, 14(1):119–139, 2005.

[20] D.B. Edelman, B.J. Baars, and A.K. Seth. Identifying the hallmarks of consciousness in non-mammalian species. *Consciousness and Cognition*, 14(1):169–187, 2005.

[21] S. Laureys. The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends Cogn Sci*, 9:556–559, 2005.

[22] G. Tononi and G.M. Edelman. Consciousness and complexity. *Science*, 282:1846–1851, 1998.

[23] A.K. Seth, E. Izhikevich, G.N. Reeke, and G.M. Edelman. Theories and measures of consciousness: An extended framework. *Proceeding of the National Academy of Sciences, USA*, 103(28):10799–10804, 2006.

[24] G. Tononi. An information integration theory of consciousness. *BMC Neuroscience*, 5(1):42, 2004.

[25] M.S. Gazzaniga. Forty-five years of split-brain research and still going strong. *Nat Rev Neurosci*, 6:653–659, 2005.

[26] H. Pashler. Dual-task interference in simple tasks: data and theory. *Psychol Bull*, 116:220–244, 1994.

[27] S. Zeki. A century of cerebral achromatopsia. *Brain*, 113 (Pt 6):1721–1777, 1990.

[28] G.M. Edelman and G. Tononi. *A universe of consciousness: How matter becomes imagination*. Basic Books, New York, 2000.

[29] G. Tononi, A.R. McIntosh, D.P. Russell, and G.M. Edelman. Functional clustering: identifying strongly interactive brain regions in neuroimaging data. *Neuroimage*, 7:133–149, 1998.

[30] G. Tononi, O. Sporns, and G.M. Edelman. Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex*, 2(4):31–35, 1992.

[31] A.K. Seth, J.L. McKinstry, G.M. Edelman, and J.L. Krichmar. Visual binding through reentrant connectivity and dynamic synchronization in a brain-based device. *Cerebral Cortex*, 14:1185–99, 2004.

[32] G.M. Edelman. *Neural Darwinism*. Basic Books, New York, 1987.

[33] G.M. Edelman. Selection and reentrant signaling in higher brain function. *Neuron*, 10:115–125, 1993.

[34] K. J. Friston, G. Tononi, G. N. Reeke, O. Sporns, and G. M. Edelman. Value-dependent selection in the brain: Simulation in a synthetic neural model. *Neuroscience*, 59:229–243, 1994.

[35] G. Tononi, O. Sporns, and G.M. Edelman. Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Science (USA)*, 96:3257–3262, 1999.

[36] G.M. Edelman and J. Gally. Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences, USA*, 98(24):13763–13768, 2001.

[37] J. Kim. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. MIT Press/Bradford Books, Cambridge, MA, 1998.

[38] D. Wegner, editor. *The illusion of conscious will*. MIT Press, Cambridge, MA, 2003.

[39] E. Thompson and F. Varela. Radical embodiment: Neural dynamics and consciousness. *Trends Cogn Sci*, 5:418–425, 2001.

[40] G. Tononi, O. Sporns, and G.M. Edelman. A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Science (USA)*, 91:5033–5037, 1994.

[41] W. Vanduffel, B.R. Payne, S.G. Lomber, and G.A. Orban. Functional impact of cerebral connections. *Proceedings of the National Academy of Science (USA)*, 94:7617–7620, 1997.

[42] A. Papoulis and S.U. Pillai. *Probability, random variables, and stochastic processes*. McGraw-Hill, New York, NY, 2002. 4th edition.

[43] D.S. Jones. *Elementary information theory*. Clarendon Press, 1979.

[44] W.J. McGill. Multivariate information transmission. *IEEE Trans Inform Theory*, 4:93–111, 1954.

[45] M. de Lucia, M. Bottaccio, M. Montuori, and L. Pietronero. A topological approach to neural complexity. *Phys. Rev. E*, 71:016114, 2004.

[46] G. Tononi, G.M. Edelman, and O. Sporns. Complexity and coherency: Integrating information in the brain. *Trends in Cognitive Science*, 2:474–484, 1998.

[47] S. L. Bressler. Large-scale cortical networks and cognition. *Brain Res Brain Res Rev*, 20:288–304, 1995.

[48] K. Friston. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annu Rev Neurosci*, 25:221–250, 2002.

[49] O. Sporns, G. Tononi, and G.M. Edelman. Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex*, 10:127–

141, 2000.

[50] O. Sporns and G. Tononi. Classes of network connectivity and dynamics. *Complexity*, 7(1):28–38, 2002.

[51] A.K. Seth and G.M. Edelman. Theoretical neuroanatomy: Analyzing the structure, dynamics and function of neuronal networks. In E. Ben Naim, H. Fraunfelder, and Z. Toroczkai, editors, *Complex networks*, Lecture Notes in Physics, pages 487–518. Springer-Verlag, Berlin, 2004.

[52] O. Sporns, D. Chialvo, M. Kaiser, and C. Hilgetag. Organization, development and function of complex brain networks. *Trends Cogn Sci*, 8:418–425, 2004.

[53] C.L. Buckley and S. Bullock. Spatial embedding and complexity: The small-world is not enough. In F. Almeida e Costa, editor, *Proceedings of the Ninth European Conference on Artificial Life*, pages 986–995. Springer-Verlag, 2007.

[54] M. Mitchell. *An introduction to genetic algorithms*. MIT Press, Cambridge, MA, 1997.

[55] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small world' networks. *Nature*, 393:440–442, 1998.

[56] O. Sporns. Complex neural dynamics. In V.K. Jirsa and J.A.S. Kelso, editors, *Coordination dynamics: Issues and trends*, pages 197–215. Springer Verlag, Berlin, 2004.

[57] O. Sporns and R. Kötter. Motifs in brain networks. *PLoS Biol*, 2:e369–e369, 2004.

[58] O. Sporns. Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems*, 85:55–64, 2006.

[59] A.K. Seth and G.M. Edelman. Environment and behavior influence the complexity of evolved neural networks. *Adaptive Behavior*, 12:5–21, 2004.

[60] O. Sporns and M. Lungarella. Evolving coordinated behavior by maximizing information structure. In L. Rocha, L. Yaeger, M.A. Bedau, D. Floreano, R.L. Goldstone, and A. Vespignani, editors, *Proceedings of the 10th European Conference on Artificial Life*, pages 323–330, Cambridge, MA, 2006. MIT Press.

[61] M. Lungarella and O. Sporns. Mapping information flow in sensorimotor networks. *PLoS Comput Biol*, 2:e144–e144, 2006.

[62] G. Tononi, O. Sporns, and G.M. Edelman. A complexity measure for selective matching of signals by the brain. *Proceedings of the National Academy of Science (USA)*, 93:3422–3427, 1996.

[63] G. Tononi and O. Sporns. Measuring information integration. *BMC Neuroscience*, 4(1):31, 2003.

[64] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–4, 2000.

[65] G. Werner. Perspectives on the neuroscience of cognition and consciousness. *Biosystems*, 87:82–95, 2007.

[66] A.K. Seth. Causal connectivity of evolved neural networks during behavior. *Network: Computation in Neural Systems*, 16:35–54, 2005.

[67] C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[68] A.K. Seth. Granger causality. *Scholarpedia*, page 15501, 2007.

[69] J.D. Hamilton. *Time series analysis*. Princeton University Press, Princeton, NJ, 1994.

[70] J. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77:304–13, 1982.

[71] A.K. Seth. Causal networks in simulated neural systems. *Cognitive Neurodynamics*, page xxx, 2008.

[72] A.K. Seth and G.M. Edelman. Distinguishing causal interactions in neural populations. *Neural Computation*, 19:910–933, 2007.

[73] M. Ding, Y. Chen, and S. Bressler. Granger causality: Basic theory and application to neuroscience. In S. Schelter, M. Winterhalder, and J. Timmer, editors, *Handbook of Time Series Analysis*, pages 438–460. Wiley, Wienheim, 2006.

[74] A. Zellner. *An introduction to Bayesian inference in econometrics*. Wiley, New York, 1971.

[75] N. Ancona, D. Marinazzo, and S. Stramaglia. Radial basis function approaches to nonlinear granger causality of time series. *Physical Review E*, 70:056221, 2004.

[76] Y. Chen, G. Rangarajan, J. Feng, and M. Ding. Analyzing multiple nonlinear time series with extended Granger causality. *Physics Letters A*, 324:26–35, 2004.

[77] S. Gao, A.K. Seth, K. Kendrick, and J. Feng. Partial granger causality: Eliminating exogenous input. *submitted*.

[78] J.E. Bogen. On the neurophysiology of consciousness: I. An overview. *Conscious Cogn*, 4:52–62, 1995.

[79] B. Kolb and I.Q. Whishaw. *Fundamentals of human neuropsychology*. W.H. Freeman, New York, NY, 4th edition, 1996.

[80] B.J. Baars. *A cognitive theory of consciousness*. Cambridge University Press, New York, NY, 1988.

[81] B.J. Baars. The conscious access hypothesis: Origins and recent evidence. *Trends Cogn Sci*, 6:47–52, 2002.

[82] H.C. Lau and R.E. Passingham. Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc Natl Acad Sci U S A*, 103:18763–18768, 2006.

[83] S. Dehaene, L. Naccache, L. Cohen, D. L. Bihan, J. F. Mangin, J. B. Poline, and D. Rivière. Cerebral mechanisms of word masking and unconscious repetition priming. *Nat Neurosci*, 4:752–758, 2001.

[84] A. Floyer-Lea and P. M. Matthews. Changing brain networks for visuomotor control with increased movement automaticity. *J Neurophysiol*, 92:2405–2412, 2004.

[85] R. Blake and N. Logothetis. Visual competition. *Nat Rev Neurosci*, 3:13–21, 2002.

[86] G. Tononi, R. Srinivasan, D.P. Russell, and G.M. Edelman. Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proceedings of the National Academy of Science (USA)*, 95:3198–3203, 1998.

[87] R. Srinivasan, D.P. Russell, G.M. Edelman, and G. Tononi. Increased synchronization of magnetic responses during conscious perception. *Journal of Neuroscience*, 19:5435–5448, 1999.

[88] R. B. Silberstein, M. A. Schier, A. Pipingas, J. Ciorciari, S. R. Wood, and D. G. Simpson. Steady-state visually evoked potential topography associated with a visual vigilance task. *Brain Topogr*, 3:337–347, 1990.

[89] D. Cosmelli, O. David, J.-P. Lachaux, J. Martinerie, L. Garnero, B. Renault, and F. Varela. Waves of consciousness: Ongoing cortical patterns during binocular rivalry. *Neuroimage*, 23:128–140, 2004.

[90] P. Nunez and R. Srinivasan. A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness. *Clin Neurophysiol*, 117:2424–2435, 2006.

[91] Y. Chen, A.K. Seth, J.A. Gally, and G.M. Edelman. The power of human brain magnetoencephalographic signals can be modulated up or down by changes in an attentive visual task. *Proc Natl Acad Sci U S A*, 100:3501–3506, 2003.

[92] R. Srinivasan. Internal and external neural synchronization during conscious perception. *Int J Bifurcat Chaos*, 14:825–842, 2004.

[93] G.A. Mashour. Consciousness unbound: Toward a paradigm of general anesthesia. *Anesthesiology*, 100:428–433, 2004.

[94] E. R. John, L. S. Prichep, W. Kox, P. Valdés-Sosa, J. Bosch-Bayard, E. Aubert, M. Tom, F. di Michele, L. D. Gugino, and F. diMichele. Invariant reversible QEEG effects of anesthetics. *Conscious Cogn*, 10:165–183, 2001.

[95] M. Massimini, F. Ferrarelli, R. Huber, S.K. Esser, H. Singh, and G. Tononi. Breakdown of cortical effective connectivity during sleep. *Science*, 309:2228–2232, 2005.

[96] C. J. Stam, B. F. Jones, G. Nolte, M. Breakspear, and Ph Scheltens. Small-world networks and functional connectivity in Alzheimer's disease. *Cereb Cortex*, 17:92–99, 2007.

[97] K. E. Stephan, C. C. Hilgetag, G. A. Burns, M. A. O'Neill, M. P. Young, and R. Kötter. Computational analysis of functional connectivity between areas of primate cerebral cortex. *Philos Trans R Soc Lond B Biol Sci*, 355:111–126, 2000.

[98] A. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.
[99] W.R. Ashby. *Design for a brain: The origin of adaptive behaviour*. Chapman Hall, London, 1952.
[100] A. Katchalsky, V. Rowland, and B. Hubermann. *Neurosci. Res. Prog. Bull.*, 12, 1974.
[101] H. Haken. *Brain dynamics*. Springer, New York, NY, 2002.
[102] W.J. Freeman. *Mass action in the nervous system*. Academic Press, New York, NY, 1975.
[103] W.J. Freeman. A field-theoretic approach to understanding scale-free neocortical dynamics. *Biological Cybernetics*, 92(6):350–359, 2005.
[104] J.A.S. Kelso. *Dynamic patterns: The self-organisation of brain and behavior*. MIT Press, Cambridge, MA, 1995.
[105] G. Werner. Metastability, criticality and phase transitions in brain and its models. *Biosystems*, 90:496–508, 2007.
[106] E.M. Izhikevich. *Dynamical systems in neuroscience: The geometry and excitability of bursting*. MIT Press, Cambridge, MA, 2006.
[107] S. Bressler and J. Kelso. Cortical coordination dynamics and cognition. *Trends Cogn Sci*, 5:26–36, 2001.
[108] K. J. Friston. Transients, metastability, and neuronal dynamics. *Neuroimage*, 5:164–171, 1997.
[109] A. Fingelkurts and A. Fingelkurts. Making complexity simpler: Multivariability and metastability in the brain. *Int J Neurosci*, 114:843–862, 2004.
[110] B.J. Baars. Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Prog Brain Res*, 150:45–53, 2005.
[111] S. Dehaene, C. Sergent, and J.-P. Changeux. A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci U S A*, 100:8520–8525, 2003.
[112] R. Wallace. *Consciousness: A mathematical treatment of the neuronal global workspace model*. Springer, New York, NY, 2005.
[113] S. Dehaene, M. Kerszberg, and J. P. Changeux. A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci U S A*, 95:14529–14534, 1998.
[114] S. Dehaene and J.-P. Changeux. Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biol*, 3:e141–e141, 2005.
[115] B. Bollobás. *Random graphs*. Academic Press, London, 1985.
[116] C. von der Malsburg. Binding in models of perception and brain function. *Curr Opin Neurobiol*, 5:520–526, 1995.
[117] W. Singer and C. M. Gray. Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci*, 18:555–586, 1995.
[118] P. Fries, J. H. Reynolds, A. E. Rorie, and R. Desimone. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291:1560–1563, 2001.
[119] P. N. Steinmetz, A. Roy, P. J. Fitzgerald, S. S. Hsiao, K. O. Johnson, and E. Niebur. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404:187–190, 2000.
[120] A. K. Engel, P. Fries, P. König, M. Brecht, and W. Singer. Temporal binding, binocular rivalry, and consciousness. *Conscious Cogn*, 8:128–151, 1999.
[121] L. Melloni, C. Molina, M. Pena, D. Torres, W. Singer, and E. Rodriguez. Synchronization of neural activity across cortical areas correlates with conscious perception. *J Neurosci*, 27:2858–2865, 2007.
[122] S. Palva, K. Linkenkaer-Hansen, R. Näätänen, and J. Palva. Early neural correlates of conscious somatosensory perception. *J Neurosci*, 25:5248–5258, 2005.
[123] K. J. Meador, P. G. Ray, J. R. Echauz, D. W. Loring, and G. J. Vachtsevanos. Gamma coherence and conscious perception. *Neurology*, 59:847–854, 2002.
[124] M. N. Shadlen and J. A. Movshon. Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron*, 24:67–77, 111, 1999.

[125] F. Crick and C. Koch. A framework for consciousness. *Nat Neurosci*, 6:119–126, 2003.

[126] D.O. Hebb. *The organization of behavior*. Wiley, New York, NY, 1949.

[127] V.A.F. Lamme. Towards a true neural stance on consciousness. *Trends Cogn Sci*, 10:494–501, 2006.

[128] D.J. Chalmers. *The conscious mind: In search of a fundamental theory*. Oxford University Press, Oxford, 1996.

[129] M.R. Bennett and P.M.S. Hacker. *Philosophical foundations of neuroscience*. Blackwell, Oxford, 2003.

[130] E. Thompson. Life and mind: A tribute to Francisco Varela. *Phenomenology and the cognitive sciences*, 3:381–98, 2004.

[131] E. Izhikevich, J.A. Gally, and G.M. Edelman. Spike-timing dynamics of neuronal groups. *Cerebral Cortex*, 14:933–944, 2004.