Classification: Biological Sciences: Neuroscience

Theories and Measures of Consciousness: An Extended Framework

Anil K. Seth[1], Eugene M. Izhikevich[1], George N. Reeke[1,2], Gerald M. Edelman[1]

[1]The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121

[2]The Rockefeller University, 1230 York Avenue, New York, NY 10021

Corresponding author:

edelman@nsi.edu, tel: 858 626 2000, fax 858 626 2099

Manuscript information:

23 pages of text including title, abstract, and references

39,652 characters (including spaces)

1680 (table allowance)

360 (equation allowance)

41,692 total characters

Abbreviations:

TNGS: Theory of Neuronal Group Selection

Abstract

A recent theoretical emphasis on complex interactions within neural systems underlying consciousness has been accompanied by proposals for the quantitative characterization of these interactions. Here, we distinguish key aspects of consciousness that are amenable to quantitative measurement from those that are not. We carry out a formal analysis of the strengths and limitations of three quantitative measures of dynamical complexity in the neural systems underlying consciousness: neural complexity, information integration, and causal density. We find that no single measure fully captures the multidimensional complexity of these systems and all have practical limitations. Our analysis suggests guidelines for the specification of alternative measures which, in combination, may improve the quantitative characterization of conscious neural systems. Given that some aspects of consciousness are likely to resist quantification altogether, we conclude that a satisfactory theory is likely to be one that combines both qualitative and quantitative elements.

Introduction

Any scientific study of consciousness is based on the premise that phenomenal experience is entailed by neuronal activity in the brain. Given this premise, an adequate theory of consciousness must be consistent with physics and with evolutionary principles. Non-physical or dualistic forces or processes must be excluded, and neural mechanisms of consciousness must emerge ontogenetically and provide adaptive advantage to a species via the ongoing exchange of signals among brains, bodies, and environments. Ideally, a theory of consciousness should propose neural mechanisms that account for its various features, which range from the multimodal characteristics of conscious scenes to the emergence of a first-person perspective (1, 2). An adequate theory should also consider whether certain of these features are susceptible to a quantitative analysis. In this regard, a fundamental property of conscious scenes is that they are both *differentiated* (reflecting the discriminatory capability of consciousness; i.e., every conscious scene is one among a vast repertoire of different possible conscious scenes) and also *integrated* (reflecting the unity of conscious experience; every conscious scene is experienced 'all of a piece') (1, 3). In this paper we summarize a theoretical framework (1) provided by the theory of neuronal group selection (TNGS), which is consistent with these requirements (1, 4-6). We then extend this framework by considering the strengths and limitations of several formal measures that have been proposed to characterize the balance between differentiation and integration in the complex neuronal dynamics responsible for consciousness (1, 3, 7). We refer to this balance as the 'relevant complexity' of the system.

According to the TNGS, the brain is a selectional system and not an instructional system like a computer. During development and behavior, vast numbers of variant neuronal circuits are generated. These constitute complex repertoires from which circuits shaped by the constraints of value systems are selected to assure adaptive behavior of the organism. In this context, 'value' refers to the positive or negative salience of an event for the organism, as determined by evolution and learning. Value is mediated by diffuse ascending neural pathways originating, for example, in dopaminergic, catecholaminergic,

and cholinergic brainstem nuclei. Spatiotemporal coordination of the neural activity underlying these selectional events is achieved mainly by a process of reentry. Reentry is the dynamic recursive exchange of signals across massively parallel axonal systems that reciprocally link maps and nuclei in the brain.

The TNGS proposes that consciousness is entailed by extensive reentrant interactions among neuronal populations in the thalamocortical system, the so-called dynamic core (1, 3, 7-10). 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 categorization 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. A key claim of the TNGS is that conscious qualia *are* these high-dimensional discriminations (1, 9). Just as conscious scenes are both differentiated and integrated at the phenomenal level to yield high-dimensional discriminations, so also are the reentrant dynamics of their underlying neural mechanisms differentiated and integrated. Useful measures of the relevant complexity of the neural systems underlying consciousness should therefore reflect this dynamic balance in the activity of the dynamic core.

To be useful, a quantitative measure should satisfy several constraints. Inasmuch as conscious experience is engendered by physical neuronal operations within the brain, a suitable measure should reflect the fact that consciousness is a dynamic *process* (11), not a *thing* or a *capacity*. This point is particularly important in light of the observation that conscious systems are embodied and bodies are embedded in and act within environments. Conscious scenes arise ultimately from transactions between organisms and environments, and these transactions are fundamentally processes. 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.

The ability to assess quantifiable aspects of consciousness at the neural level without first person report would be useful for the assessment of depth of anesthesia (12) as well as in the analysis of various neurological and psychiatric disorders (13). Application of quantitative measures may also contribute to comparative studies of consciousness. The attribution of conscious states to non-human animals is made difficult by their inability verbally to report the contents of their putative consciousness (14). A quantitative measure of relevant complexity might provide one criterion for assessing the relative degree of consciousness in such non-human animals (15).

Obviously, the quantitative characterization of relevant complexity can only constitute 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 (1, 14), several of which do not appear to be quantifiable by a single measure (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?

Measuring relevant complexity

In the following, we critically examine three proposed measures of the relevant complexity of conscious neural systems: neural complexity, C_N, information integration, Φ, and a new measure, causal density, *cd* (16-19). To our knowledge, these are the only extant measures that explicitly attempt to quantify the balance between integration and differentiation exhibited by a neural system. While these and related measures might also be applicable to non-neural systems, we are concerned here in the main with neural systems only. In our analysis of these measures, we investigate how well the constraints of process-orientation, causality, and computability are satisfied. We exclude from detailed consideration properties of neuronal dynamics, such as synchrony (20), for which explicit measures that can be associated with relevant complexity have not been proposed. Moreover, we do not consider several theoretical perspectives that share many common features with the dynamic core hypothesis (1), but which are not explicitly concerned with quantitative measures of complex dynamics. They include the notions of 'coalitions' of neurons (21), global negatively entropic brain states (22), the 'global workspace' (23), and association of perceptual events with the coalescence of a 'macroscopic pool' of mesoscopic 'wave packets' of neural activity (24).

Neural Complexity

Neural complexity expresses the extent to which a system is both dynamically segregated, so that small subsets of the system tend to behave independently, and dynamically integrated, so that large subsets of the system tend to behave coherently (3, 7, 16). A practical algorithm for the computation of neural complexity is provided in (16) and also in Supplementary Information S1. In brief, the neural complexity C_N of a system X composed of *n* elements is equal to the sum of the average mutual information across all bipartitions of the system (16). The mutual information between two subsets A and B, defined by a single bipartition, measures the uncertainty about A that is accounted for by the state of B. It is calculated as MI(A;B) = H(A) + H(B) – H(AB), where H(.) is the informational entropy, i.e., the overall degree of statistical independence. Under

Gaussian assumptions, the entropy of the system, H(X), or the entropy of any subset of the system, can be calculated analytically from the covariance matrix COV(X) relating the responses of the elements of the system.

The covariance matrix COV(X) can in turn be calculated analytically from the system's connectivity matrix $C_{ij}(X)$, assuming linear system dynamics and activation of network elements by uncorrelated noise (16). Alternatively, COV(X) can be derived empirically on the basis of the recorded activity of a network over a specific time period. In this case, C_N reflects the explicit exchange of signals that takes place either within the isolated system or, in a behaving system, during interaction with an external environment as an embedded and embodied neural network (25). The concept of neural complexity has been extended to characterize the selectional responses of neural systems to inputs in terms of 'matching' complexity, which is calculated as the total neural complexity of a neural system X when the input is present, minus the intrinsic complexity of X and minus the complexity that is directly attributable to the input (26).

Precise calculation of C_N requires the evaluation of mutual information across all possible bipartitions, which can become computationally prohibitive for large systems. There is, however, a tractable approximation to C_N which, instead of considering all possible bipartitions of a system, considers only those that divide the system into sets comprising one single element and all the remaining elements [see (27) and Supplementary Information S1]. A disadvantage of C_N and its approximation is that they do not reflect causal interactions. This is so because C_N is based on mutual information, which is a symmetric quantity.

Information Integration, Φ

This measure has been proposed as a way to quantify the total amount of information that a conscious system can integrate (18). The theory in which Φ is proposed as the central element, the information integration theory of consciousness [IITC, (18)], makes the claim that consciousness corresponds to the capacity of a system

to integrate information, and that Φ measures this capacity: "experience, that is, information integration, is a fundamental quantity, just as mass, charge or energy are. It follows that any physical system has subjective experience to the extent that it is capable of integrating information" [(18), p.19]. Φ is defined in (18) as the 'effective information' across the informational 'weakest link' of a system, the so-called 'minimum information bipartition' [MIB; see (17, 18) and Supplementary Information S2]. Effective information is calculated as the mutual information across a partition in the case where outputs from one subset have maximum entropy, and the MIB is that partition of the system for which the effective information is lowest.

Inasmuch as Φ was explicitly formulated to measure consciousness as a capacity as opposed to a process, two features are critical: (1) determining Φ depends on replacing the outputs of all possible subsets of a system with uncorrelated noise, so that each set of outputs has maximum entropy (i.e., reflecting *all possible* activity patterns), and (2) the effective information across the majority of partitions is significant only insofar as it helps determine which partition is the MIB; the value of Φ depends only on the effective information across the MIB. The focus on capacity leads to the counterintuitive prediction (18) that a brain with high value of Φ but displaying no activity at all would be conscious.

Unlike C_N, Φ reflects causal interactions. This is so because Φ is based on effective information, which is a directional version of mutual information that relies on the replacement of the outputs of different subsets of the studied system with maximum entropy signals. However, Φ cannot be measured for any non-trivial real-world system, for two reasons. First, it is infeasible to replace the outputs of arbitrary subsets of complex real neural systems with uncorrelated noise. Second, the evaluation of Φ requires the calculation of effective information across each bipartition of a system, and there is a factorial growth in the number of partitions that must be examined as the size of the network increases, i.e., as with C_N, the number of partitions grows approximately as n^n for networks of size n. Although the possibility of confronting this issue has been

discussed (17), absent an effective approximation, the evaluation of Φ is computationally infeasible for large networks.

In contrast to neural complexity C_N and causal density *cd* (19), Φ has been proposed as an adequate measure of the "quantity" of consciousness generated by a system, such that systems with sufficiently high values of Φ would necessarily be conscious (18). It is therefore critical for the IITC that high values of Φ should not be obtained from arbitrary non-conscious systems. However, we here show analytically that, even for a trivially simple network, Φ may grow without bound as a function of network size.

Consider a fully-connected Hopfield-type network (28) with synaptic weights from the *j*-th neuron ($j = 1,..,n$) to the *i*-th neuron ($i = 1,..,n$) defined by $C_{ij}=2^j$ so that network activity is updated according to

$$x_i(t+1) = \sum_{j=1}^{n} 2^j f(x_j(t)), \quad \text{where } f(x) = \begin{cases} -1 \; if \; x < 0 \\ +1 \; if \; x \geq 0 \end{cases}, \tag{1}$$

and where each variable $x_i(t)$ describes the (integer value) state of the *i*-th neuron at time *t*. Consider now a (*k*, *n-k*) bipartition, A|B, of the network, i.e., a subset A with *k* neurons and a subset B with *n-k* neurons. The effective information EI(A→B) is given by EI(A→B) = H(A) + H(B) – H(AB) under conditions in which outputs from A are replaced by uncorrelated noise, as specified by the definition of Φ (18). As we show in detail in Supplementary Information S2, for the network (1),

$$\text{EI}(A \rightarrow B) = k \text{ bits.} \tag{2}$$

Similarly, EI(B→A) is equal to *n-k* bits, which implies that

$$\text{EI}(A \leftrightarrow B) = \text{EI}(A \rightarrow B) + \text{EI}(B \rightarrow A) = n \text{ bits.} \tag{3}$$

Since equation (3) does not depend on k, the effective information across every bipartition is the same. It is easy to check that the effective information across every bipartition of a subset of m neurons ($m<n$) is equal to m (see Supplementary Information S2). Therefore, the information integration value for the complete network is given by

$$\Phi = n \text{ bits.} \qquad (4)$$

The above result implies that for any value of Φ associated with a presumably conscious neural system, there exists a simple Hopfield-type network that has an equivalent or greater Φ, and that would, lead by a key assumption of the IITC (18), to the conclusion that this network is conscious.

Given this assumption (18), it also seems critical for the applicability of the IITC that any measured value of Φ *not* depend on arbitrary choices made by an observer. However, any quantitative measure of relevant complexity, including C_N, Φ, and cd, will vary according to the variables chosen to characterize the system. With regard to both Φ and C_N, any measured value involves the calculation of informational entropy, which requires the identification of a repertoire of states to which probabilities of occurrence can be assigned. For complex neural systems, the identification of such a repertoire depends on arbitrary choices for the reason that such systems can be described by many different variables, such as transmembrane potentials, action potentials, and local field potentials. The repertoire of states corresponding to each variable, or to any combination of variables will, in general, be different, and therefore the corresponding values for entropy will also be different. Furthermore, the variables describing complex neural systems are usually continuous – even an action potential is a continuous event if the voltage spike is plotted on a sub-millisecond time scale – which implies a further dependency on the observer in the specification of the units in which a given variable is measured. As with the choice of variables, this dependency on measurement units applies equally to C_N, Φ, and cd . In Supplementary Information S2 we show that a simple continuous system consisting of two coupled oscillators can generate an arbitrary, even infinite, value for Φ depending on the measurement units selected by the observer.

Causal Density

A balance between dynamical integration and differentiation is likely to involve dense networks of causal interactions among neuronal elements. Causal density (*cd*) is a novel measure of causal interactivity that captures dynamical heterogeneity among network elements (differentiation) as well as their global dynamical integration (see (19) and Supplementary Information S3). Specifically, *cd* is a measure of the fraction of interactions among neuronal elements that are causally significant. It can be calculated by applying 'Granger causality' (29), 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. In practice, Granger causality is tested in the context of multivariate linear regression models relating the activities of the elements of the system (30).

To illustrate Granger causality, 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:

$$x_1(t) = \sum_{j=1}^{p} A_{11,j}\, x_1(t-j) + \sum_{j=1}^{p} A_{12,j}\, x_2(t-j) + E_1(t)$$
$$x_2(t) = \sum_{j=1}^{p} A_{21,j}\, x_1(t-j) + \sum_{j=1}^{p} A_{22,j}\, x_2(t-j) + E_2(t) \qquad (5)$$

where p is the maximum number of lagged observations included in the model (the model order, $p<T$), the matrix A contains the coefficients of the model [i.e., the contributions of each lagged observation to the predicted values of $x_1(t)$ and $x_2(t)$], and E_1 and E_2 are residuals (prediction errors) for each time series. If the variance of E_1 (or E_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) *Granger-causes* x_1 (or x_2). In other words, x_2 Granger-causes x_1 if the coefficients in A_{12} are jointly significantly different from zero. This can be tested by performing an F-test of the null hypothesis that $A_{12} = 0$, given assumptions of covariance stationarity on x_1 and x_2. The magnitude of a Granger causality interaction can be

estimated by the logarithm of the corresponding F-statistic (31). For present purposes, it is important to note that this concept can be readily extended to the *n* variable case, where $n > 2$, by estimating an *n* variable autoregressive model. In this case, x_2 Granger-causes x_1 if lagged observations of x_2 help predict x_1 when lagged observations of all other variables $x_{3 \ldots n}$ are also taken into account.

Following a Granger causality analysis, the causal density of a system can be calculated as

$$cd = \alpha / n(n-1) \tag{6}$$

where α is the total number of significant causal interactions and $n(n\text{-}1)$ is the total number of directed edges in a fully connected network with *n* nodes, excluding self-connections. 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).

Causal density is inherently a measure of process. It 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. Because causal density is based on a well-established statistical interpretation of causality, it incorporates causal interactions by design. We emphasize that the value of *cd* for a system depends on *all* causal interactions within the system, and not just on those interactions across a single bipartition, as is the case for Φ.

A practical problem with the determination of 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 (see equations 5). 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 (32). In neural systems, such prior constraints may be derived, for example, on the basis of known neuroanatomy or by anatomically-based clustering procedures.

From the above analyses, we conclude that while existing formal measures may have heuristic value in highlighting the need to characterize differentiation and integration in the dynamic core, 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, but is a measure of capacity not process, is infeasible to compute in neural systems of non-trivial size, and can be shown to grow without bound even for certain simple networks. *cd* reflects all causal interactions within a system and is explicitly a measure of process, but it also may be difficult to compute for large systems.

The existence of quantitative measures of relevant complexity, however preliminary they may be, raises the important issue of identifying the ranges of values that would be consistent with consciousness. As we have mentioned, all of the measures analyzed above are necessarily based on an exogenously selected repertoire of variables and of units of measurement for these variables. This dependency emphasizes the requirement that any proposed quantitative measure of consciousness be embedded in a qualitative brain theory in order to justify and inform such exogenous selections. These selections having been made, it may then become possible to define a measurement scale (33) for a proposed measure of relevant complexity by establishing a value for a known conscious system (for example, an awake human) and a value for a known non-conscious system (for example, the same human during dreamless sleep).

Dimensions of relevant complexity

In addition to analyzing whether a proposed measure satisfies the constraints of process-orientation, causality, and computability, it is important to consider further whether any one measure could be sufficient to assess the relevant complexity of a conscious neural system. We suggest that characterizing the relevant complexity of such a system will require a multidimensional analysis of transactions within the thalamocortical core. Such a multidimensional analysis is in turn likely to require the simultaneous application of multiple formal measures. For example, we can identify three distinct dimensions along which the relevant complexity of a conscious neural system is likely to vary: spatial complexity, temporal complexity, and recursive complexity. The salience of these dimensions can be seen, as described below, by their correspondence to aspects of phenomenal experience. While reference to these dimensions may not fully exhaust the relevant complexity of any conscious neural system, they may provide guidelines for the development of useful quantitative measures.

Before describing further the above dimensions of relevant complexity, it is important to distinguish the space incorporating these dimensions from the proposed concept of a multidimensional 'qualia space' (3). Qualia space is a high-dimensional space in which the axes reflect dimensions on which phenomenally experienced conscious scenes are discriminated (e.g., color, shape, smell, touch, proprioception, etc.). The concept of qualia space reflects the observation that conscious scenes consist of enormously informative discriminations among a vast repertoire of possible experiences. By contrast, the dimensions along which relevant complexity can be measured reflect the activity of the *physical* machinery, mainly in the dynamic core, which entails these phenomenal discriminations. Therefore, although each dimension of relevant complexity bears upon aspects of phenomenal experience, there is no necessary one-to-one correspondence between dimensions of relevant complexity and dimensions of qualia space.

Spatial complexity reflects the balance between integration and differentiation in the spatial domain. Not surprisingly, existing measures of relevant complexity, such as C_N, Φ, and *cd*, have focused largely on spatial complexity. The spatial structure of a conscious scene, which is most salient in the visual modality, is both unified into a Gestalt and differentiated into individual features. At the neural level, it is well established that different brain regions are specialized for different functions and that the activities of these diverse regions must be globally coordinated in order to yield coherent behavior. Thus, any theory of consciousness that involves interactions among spatially distributed and functionally segregated brain regions immediately faces the task of characterizing spatial complexity.

Temporal complexity reflects the fact that consciousness extends over time in several ways. For example, consciousness is associated with the ordering of events into complex temporal sequences. Musical and linguistic phrases, which require consciousness for their full apprehension, are prototypical examples of complex temporal sequences that cannot be reduced to simple associative chains (34, 35). Other examples include the construction of an internal historical narrative based on episodic memories, and the projection of such a narrative into a conditional future (8). Conscious effort also appears to be required for the initial learning of complex motor sequences (36), a notion supported by neuroimaging studies showing widespread cortical activation during early learning as compared to during expression of learned behavior (37-39).

Consciousness itself involves the generation of a subjective 'now' or 'remembered present' (8). Empirical studies suggest that it takes ~100 ms for sensory stimuli to be incorporated into a conscious scene (40), and that neuronal 'readiness potentials' can appear several hundred ms prior to reportable awareness of intentions to act (41). Moreover, conscious scenes subjectively attributed to a particular time can be influenced by physical events happening after this time (42). In general, the generation of each conscious scene involves the integration of ongoing signals reflecting sensation and intention with those reflecting a past history of value-dependent categorization, learning, and memory.

The above observations are consistent with the notion that conscious experiences are both differentiated over time (each temporal component is distinct), and integrated over time (time is experienced as a continuum, stretching from a definite past towards an indeterminate future). Temporal complexity therefore parallels spatial complexity in reflecting the balance between differentiation and integration, but in the temporal domain. Accordingly, with suitable modification, some measures of spatial complexity may also be applicable in the temporal domain.

The notion of recursive complexity refers to the balance between differentiation and integration across different levels of description within a system. At the neural level, brains exhibit rich organization at multiple levels of description, ranging from molecular interactions within individual synapses, to the dynamics of cortical microcircuits, to reentrant interactions among functionally segregated brain regions. The phenomenal structure of consciousness also appears to be recursive; for example, the individual features of conscious scenes are themselves Gestalts, and must therefore share organizational properties with the conscious scene as a whole.

Recursive complexity is related to modular and hierarchical structures within networks. Within brains, modular and hierarchical organization reflects constraints on the genetic encoding of brain development (43) as well as anatomical constraints such as the optimization of axonal lengths (44). Hierarchical organization may also serve functional roles, for example, in the adaptive distribution of reentrant signals (45) and in providing robustness of responses to perturbations (46). Hierarchical and compositional structures may also relate to recursive relations at the phenomenal level. An example is seen in the interactions of heterogeneous elements to form new combinations, in which whole categories of things or sequences of events can be treated as single elements in a higher-level construction based on selection. This property of compositionality has been most widely discussed in contexts concerned with syntactical characteristics of human language and logical symbol systems (47, 48).

The various dimensions of relevant complexity discussed here require different strategies for their quantitative characterization. While we have considered several presently available candidate measures of the balance between differentiation and integration in the spatial domain, measures appropriate for the analysis of neural systems in the temporal domain (49, 50) and in the recursive domain remain to be adequately specified.

Summary

Given that consciousness is a rich biological phenomenon, a satisfactory neural theory of consciousness must avoid reductionistic excess. Excessive reductionism (51) can be revealed by improper reification, for example by converting a dynamic process into a static entity. It may propose the arbitrary agglomeration of different aspects of a system into a single common character, and it may involve the improper quantification of such an arbitrarily chosen character. According to these criteria, any theory that identifies consciousness with a single measure is likely to be excessively reductionistic, and as a result, limited in its scope. We suggest that the development and simultaneous application of multiple quantitative measures would more appropriately characterize the relevant complexity of the neural systems underlying consciousness. Even so, some aspects of consciousness are likely to resist quantification altogether. An adequate theory is therefore likely to be one that consists of a combination of qualitative and quantitative elements. The TNGS has been proposed with this in mind. It will be greatly enhanced when practically calculable multidimensional measures of relevant complexity are formulated.

Acknowledgements

We thank Drs. John Iversen, Joseph Gally, Luis Bettencourt, Robert Kozma, and Botond Szmatáry for useful discussions. Financial support was provided by the Neurosciences Research Foundation.

References:

1. Edelman, G. M. (2003) *Proc Natl Acad Sci U S A* **100,** 5520-4.
2. Metzinger, T. (2003) *Being No-One* (MIT Press, Cambridge, MA).
3. Edelman, G. M. & Tononi, G. (2000) *A universe of consciousness : how matter becomes imagination* (Basic Books, New York, NY).
4. Edelman, G. M. (1978) in *The Mindful Brain*, eds. Edelman, G. M. & Mountcastle, V. B. (MIT Press, Cambridge, MA).
5. Edelman, G. M. (1987) *Neural Darwinism: The Theory of Neuronal Group Selection* (Basic Books, Inc., New York).
6. Edelman, G. M. (1993) *Neuron* **10,** 115-25.
7. Tononi, G. & Edelman, G. M. (1998) *Science* **282,** 1846-51.
8. Edelman, G. M. (1989) *The remembered present* (Basic Books, New York, NY).
9. Edelman, G. M. (2004) *Wider than the sky: The phenomenal gift of consciousness* (Yale University Press).
10. Seth, A. K. & Baars, B. J. (2005) *Consciousness and Cognition* **14,** 140-168.
11. James, W. (1904) *Journal of Philosophy, Psychology, and Scientific Methods* **1,** 477-491.
12. Smith, W. D., Dutton, R. C. & Smith, N. T. (1996) *Anesthesiology* **84,** 38-51.
13. Laureys, S. (2005) in *Progress in Brain Research* (Elsevier Science, Vol. 150.
14. Seth, A. K., Baars, B. J. & Edelman, D. B. (2005) *Consciousness and Cognition* **14,** 119-139.
15. Edelman, D. B., Baars, B. J. & Seth, A. K. (2005) *Consciousness and Cognition* **14,** 169-187.
16. Tononi, G., Sporns, O. & Edelman, G. M. (1994) *Proc Natl Acad Sci U S A* **91,** 5033-7.
17. Tononi, G. & Sporns, O. (2003) *BMC Neurosci* **4,** 31.
18. Tononi, G. (2004) *BMC Neurosci* **5,** 42.
19. Seth, A. K. (2005) *Network: Computation in Neural Systems* **16,** 35-55.
20. Engel, A. K. & Singer, W. (2001) *Trends Cogn Sci* **5,** 16-25.
21. Crick, F. & Koch, C. (2003) *Nature Neuroscience* **6,** 119-126.

22. John, E. R. (2001) *Consciousness and Cognition* **10,** 184-213.
23. Baars, B. J. (1988) *A cognitive theory of consciousness* (Cambridge University Press, New York, NY).
24. Freeman, W. J. (2003) *Journal of Integrative Neuroscience* **2,** 3-30.
25. Seth, A. K. & Edelman, G. M. (2004) *Adaptive Behavior* **12,** 5-20.
26. Tononi, G., Sporns, O. & Edelman, G. M. (1996) *Proc Natl Acad Sci U S A* **93,** 3422-7.
27. Tononi, G., Edelman, G. M. & Sporns, O. (1998) *Trends Cogn Sci* **2,** 474-484.
28. Hopfield, J. J. (1982) *Proc Natl Acad Sci U S A* **79,** 2554-8.
29. Granger, C. W. J. (1969) *Econometrica* **37,** 424-438.
30. Hamilton, J. D. (1994) *Time series analysis* (Princeton University Press, Princeton, NJ).
31. Geweke, J. (1982) *Journal of the American Statistical Association* **77,** 304-313.
32. Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics* (John Wiley & Sons, New York).
33. Stevens, S. S. (1946) *Science* **103,** 677-680.
34. Van Valin, R. D. (2001) *An Introduction to Syntax* (Cambridge University Press, Cambridge).
35. Lerdahl, F. & Jackendoff, R. (1983) *A Generative Theory of Tonal Music* (MIT Press, Cambridge, MA).
36. Schneider, W. & Shiffrin, R. M. (1977) *Psychological Review* **84,** 1-66.
37. Jueptner, M., Stephan, K. M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. & Passingham, R. E. (1997) *J Neurophysiol* **77,** 1313-24.
38. Jenkins, I. H., Brooks, D. J., Nixon, P. D., Frackowiak, R. S. & Passingham, R. E. (1994) *J Neurosci* **14,** 3775-90.
39. Haier, R. J., Siegel, B. V., Jr., MacLachlan, A., Soderling, E., Lottenberg, S. & Buchsbaum, M. S. (1992) *Brain Res* **570,** 134-43.
40. Libet, B., Alberts, W. W., Wright, E. W., Jr. & Feinstein, B. (1967) *Science* **158,** 1597-600.
41. Libet, B. (1982) *Hum Neurobiol* **1,** 235-42.

42. Bachmann, T. (1993) *Psychophysiology of Visual Masking* (Nova Science, Commack, NY).
43. Geary, D. C. & Huffman, K. J. (2002) *Psychol Bull* **128,** 667-98.
44. Chklovskii, D. B., Schikorski, T. & Stevens, C. F. (2002) *Neuron* **34,** 341-7.
45. Seth, A. K., McKinstry, J. L., Edelman, G. M. & Krichmar, J. L. (2004) *Cerebral Cortex* **14,** 1185-99.
46. Variano, E. A., McCoy, J. H. & Lipson, H. (2004) *Physical Review Letters* **92,** 187701-4.
47. Pinker, S. (1994) *The Language Instinct: How the Mind Creates Language* (William Morrow).
48. Fodor, J. A. & Pylyshyn, Z. (1998) in *Connections and Symbols,* eds. Pinker, S. & Mehler, J. (MIT Press, Cambridge, MA).
49. Costa, M., Goldberger, A. L. & Peng, C. K. (2002) *Phys Rev Lett* **89,** 068102.
50. Rajkovic, M. (2004) *Physica A* **340,** 327-333.
51. Rose, S. (1997) *Lifelines: Biology, freedom, determinism* (Penguin).

1	Consciousness is accompanied by irregular, low-amplitude, fast (12-70Hz) electrical brain activity.
2	Consciousness is associated with activity within the thalamocortical complex (the 'dynamic core'), which is 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. According to the TNGS, qualia are the discriminations entailed by the underlying neural activity.
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 developed.
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.

Table 1. Thirteen features of consciousness that require theoretical explanation. Items 1 through 6 are in one degree or another susceptible to characterization by quantitative measurement. Items 7 through 13 are more readily understood through logical and qualitative analyses.